

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocelebration Ltd.

OM protein - protein search, using sw model

Run on: May 4, 2006, 03:17:19 ; Search time 163.333 Seconds
(without alignments)
64.793 Million cell updates/sec

Title: US-10-799-326-78

Perfect score: 58

Sequence: 1 DGXVIRATYXXHFXFT 15

Scoring table: BLOSUM62DX

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : UniProt 05.80:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	58	100.0	137	2	06XUH3_9CYAN
2	58	100.0	135	1	DPO3A_SYNY3
3	57	98.3	503	2	06XUH7_9NOST
4	57	98.3	870	2	05N0T2_SYNP6
5	54	93.1	877	2	08YR72_ANASP
6	46	79.3	122	2	06W5B8_SYNVU
7	46	79.3	613	2	023366_CABEL
8	46	79.3	680	2	0623A3_CABER
9	46	79.3	712	2	06XUH5_ALPHA
10	46	79.3	872	2	08DHA3_SYNEL
11	46	79.3	879	1	DNAB_ANASP
12	46	79.3	941	1	DNAB_RHOMR
13	46	79.3	1437	1	RPCS_PVRAB
14	45	77.6	872	1	DNAB_SYNY3
15	45	77.6	897	2	0906L0_BUNYW
16	44	75.9	873	2	04PRJ9_CYLFU
17	43	74.1	320	2	06LHB1_CABER
18	43	74.1	1390	2	05JHB6_PVRKO
19	42	72.4	216	2	0670G4_SYNTM
20	42	72.4	303	2	05RIK9_ORYSA
21	42	72.4	323	2	05DM17_BP75
22	42	72.4	331	2	096YR4_SULFO
23	42	72.4	434	2	07UVI1_MYCBO
24	42	72.4	434	2	08VVK1_MYCBO
25	42	72.4	457	2	06LNO7_PPROPR
26	42	72.4	550	1	FLIC_SHIFT
27	42	72.4	550	1	FLIC_SHIFT
28	42	72.4	550	2	0842D3_ECOLI
29	42	72.4	551	2	08RUA8_ORYSA
30	42	72.4	562	2	05YFA2_MAIZE
31	42	72.4	563	2	08LSK5_GOSHI

32	42	72.4	568	2	066YTB_LYCES	066YTB_LYCES
33	42	72.4	589	2	06XOM3_DICTDI	06XOM3_DICTDI
34	42	72.4	589	2	05SG10_DICTDI	05SG10_DICTDI
35	42	72.4	698	2	08XSG8_RALSO	08XSG8_RALSO
36	42	72.4	814	2	002634_HEMPU	002634_HEMPU
37	42	72.4	867	1	SYA_FUSNN	SYA_FUSNN
38	42	72.4	893	1	NIA_LEPMC	NIA_LEPMC
39	42	72.4	911	2	051ZHA_MAGGR	051ZHA_MAGGR
40	42	72.4	939	2	0766M7_NECCR	0766M7_NECCR
41	42	72.4	1075	2	08A4E4_BACTN	08A4E4_BACTN
42	42	72.4	1108	2	058310_PYRHO	058310_PYRHO
43	41	70.7	228	2	04ROG8_TETNG	04ROG8_TETNG
44	41	70.7	236	2	098DS3_RH1LO	098DS3_RH1LO
45	41	70.7	248	2	05QWY2_IDILO	05QWY2_IDILO
46	41	70.7	320	2	065GB9_BACLD	065GB9_BACLD
47	41	70.7	431	2	082S13_NITEU	082S13_NITEU
48	41	70.7	474	2	0934G9_9BACI	0934G9_9BACI
49	41	70.7	493	2	06SEYO_BACLD	06SEYO_BACLD
50	41	70.7	644	2	04HTE8_CAMUP	04HTE8_CAMUP
51	41	70.7	647	2	05HV14_CAMJR	05HV14_CAMJR
52	41	70.7	647	2	09PH10_CAMJE	09PH10_CAMJE
53	41	70.7	651	2	04HKR1_CAMLA	04HKR1_CAMLA
54	41	70.7	661	2	07BVO5_BACSU	07BVO5_BACSU
55	41	70.7	809	2	044391_STRPU	044391_STRPU
56	41	70.7	896	2	04WRV6_ASPEU	04WRV6_ASPEU
57	41	70.7	897	2	091ZKO_BUNYW	091ZKO_BUNYW
58	41	70.7	903	2	05QU14_IDILO	05QU14_IDILO
59	41	70.7	1084	2	064173_BPSPC	064173_BPSPC
60	41	70.7	1084	2	031875_BACSU	031875_BACSU
61	41	70.7	1847	1	RPCS_METUA	RPCS_METUA
62	40	69.0	75	2	06AEP9_LEIXX	06AEP9_LEIXX
63	40	69.0	113	2	083A02_COXBU	083A02_COXBU
64	40	69.0	114	2	057459_COXBU	057459_COXBU
65	40	69.0	152	2	08C292_STRRN	08C292_STRRN
66	40	69.0	152	2	097T01_STRPN	097T01_STRPN
67	40	69.0	177	2	028120_ARCFU	028120_ARCFU
68	40	69.0	225	2	04NP05_9DEL7	04NP05_9DEL7
69	40	69.0	247	2	06LSN6_NICBE	06LSN6_NICBE
70	40	69.0	271	1	Y2Z3_RHOMR	Y2Z3_RHOMR
71	40	69.0	275	2	06N9Y3_RHOPA	06N9Y3_RHOPA
72	40	69.0	296	2	096M27_RH1LO	096M27_RH1LO
73	40	69.0	314	2	086M64_TRYCR	086M64_TRYCR
74	40	69.0	321	2	09U4A6_CABEL	09U4A6_CABEL
75	40	69.0	344	2	P92631_9SNUR	P92631_9SNUR
76	40	69.0	345	2	0644U0_9SALA	0644U0_9SALA
77	40	69.0	401	2	06WXH1_SERMA	06WXH1_SERMA
78	40	69.0	401	2	07BQ40_SHIFL	07BQ40_SHIFL
79	40	69.0	401	2	07BNI8_SALTI	07BNI8_SALTI
80	40	69.0	401	2	09S458_SALTY	09S458_SALTY
81	40	69.0	401	2	07AK87_9ZGZZ	07AK87_9ZGZZ
82	40	69.0	433	2	06DU83_XENTR	06DU83_XENTR
83	40	69.0	454	2	0602H2_METCA	0602H2_METCA
84	40	69.0	467	2	04HH83_CAMCO	04HH83_CAMCO
85	40	69.0	468	2	05KV29_GEMOJA	05KV29_GEMOJA
86	40	69.0	530	2	04ZY11_PSRSY	04ZY11_PSRSY
87	40	69.0	531	2	0889C6_PSRSM	0889C6_PSRSM
88	40	69.0	639	2	07ZVF7_BRARE	07ZVF7_BRARE
89	40	69.0	680	2	05LAV3_BACFR	05LAV3_BACFR
90	40	69.0	680	2	064RA1_BACFR	064RA1_BACFR
91	40	69.0	788	2	025343_LEIMA	025343_LEIMA
92	40	69.0	811	1	ENAN_BP1LE	ENAN_BP1LE
93	40	69.0	885	2	05YNI3_NOCFA	05YNI3_NOCFA
94	40	69.0	897	2	0997B8_BUNYW	0997B8_BUNYW
95	40	69.0	920	2	04Q8N7_LEIMA	04Q8N7_LEIMA
96	40	69.0	970	2	05VIM7_HALMA	05VIM7_HALMA
97	40	69.0	972	2	095212_9RYYP	095212_9RYYP
98	40	69.0	984	2	09GQNI_CALPA	09GQNI_CALPA
99	40	69.0	984	2	09GQNI_CALPA	09GQNI_CALPA
100	40	69.0	1041	2	08MT53_DROME	08MT53_DROME

ALIGNMENTS

```

RESULT 1
06XUH3_90CYAN PRELIMINARY; PRT; 137 AA.
ID 06XUH3_90CYAN PRELIMINARY; PRT; 137 AA.
AC 06XUH3;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE DnaE (Fragment).
GN Name=dnaE;
OS Oscillatoria limnetica str. 'Solar Lake'.
OC Bacteria; Cyanobacteria; Oscillatoriales; Oscillatoria.
ON NCBI_TaxId=262926;
RX NUCLEOTIDE SEQUENCE.
RC STRAIN=Solar Lake;
RC PubMed=14651639; DOI=10.1046/j.1365-2958.2003.03825.x;
CA Caspi J., Amltai G., Belenkiy O., Petrokovski S.;
RT "Distribution of split DnaE inteins in cyanobacteria.";
RL Mol. Microbiol. 50:1569-1577(2003).
DR EMBL; AY209007; AAP47640.1; -; Genomic DNA.
DR GO; GO:0016539; P:intein-mediated protein splicing; IEA.
DR InterPro; IPR003587; Hedgehog_hint_N.
DR InterPro; IPR006141; Intein_S.
DR SMART; SM00306; HintN; 1.
DR TIGRfams; TIGR01445; Intein_Nterm; 1.
DR PROSITE; PS50817; INTEIN_N_TER; 1.
FT NON TER 1 1
SQ SEQUENCE 137 AA; 15458 MW; 7E9C8B0B73EC0AC4 CRC64;

Query Match 100.0%; Score 58; DB 2; Length 137;
Best Local Similarity 66.7%; Pred. No. 0.061;
Matches 10; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGVVIRATYXXHFXPT 15
DB 87 DGSVIRATYKHQHPMT 101

RESULT 2
DPRO3A_SYNY3 STANDARD; PRT; 1355 AA.
ID DPRO3A_SYNY3 STANDARD; PRT; 1355 AA.
AC P74750; P73215;
DT 15-DEC-1998 (Rel. 37, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Dna polymerase III alpha subunit (EC 2.7.7.7) [Contains: Ssp dnaE
intein].
GN Name=dnaE-N; OrderedLocustNames=slr0603;
GN and
OS Name=dnaE-N; OrderedLocustNames=slr11572;
OS Synchocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
ON NCBI_TaxId=1148;
RX NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naito K., Okumura S.,
RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synchocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
RX CHARACTERIZATION OF SPLIT INTEIN.
RX MEDLINE=8936139; PubMed=8689062; DOI=10.1073/pnas.95.16.9226;
RA Wu H., Hu Z., Liu X.-O.;
RT "Protein trans-splicing by a split intein encoded in a split DnaE gene
RT of Synchocystis sp. PCC6803.";
Proc. Natl. Acad. Sci. U.S.A. 95:9226-9231(1998).

```

```

RN [3]
RP CHARACTERIZATION OF SPLIT INTEIN.
RX MEDLINE=20200384; PubMed=10734038; DOI=10.1074/jbc.275.13.9091;
RA Evans T.C. Jr., Martin D., Kolly R., Pannu D., Sun L., Ghosh I.,
RA Chen L., Benner J., Liu X.-O., Xu M.-O.;
RT "Protein trans-splicing and cyclization by a naturally split intein
RT from the dnaE gene of Synchocystis species PCC6803.";
RL J. Biol. Chem. 275:9091-9094(2000).
RN [4]
RP CHARACTERIZATION OF SPLIT INTEIN.
RX MEDLINE=21093115; PubMed=11170467; DOI=10.1021/bi001786g;
RA Martin D.D., Xu M.-O., Evans T.C. Jr.
RT "Characterization of a naturally occurring trans-splicing intein from
RT Synchocystis sp. PCC6803.";
RL Biochemistry 40:1393-1402(2001).
CC -1- FUNCTION: DNA polymerase III is a complex, multichain enzyme
CC responsible for most of the replicative synthesis in bacteria.
CC This DNA polymerase also exhibits 3' to 5' exonuclease activity.
CC The alpha chain is the DNA polymerase (By similarity).
CC -1- CATALYTIC ACTIVITY: Deoxynucleoside triphosphate + DNA(n) =
CC diphosphate + DNA(n+1).
CC -1- SUBUNIT: DNA polymerase III contains a core (composed of alpha,
CC epsilon and theta chains) that associates with a tau subunit. This
CC core dimerizes to form the POLIII' complex. POLIII' associates
CC with the gamma complex (composed of gamma, delta, delta', psi and
CC chi chains) and with the beta chain to form the complete DNA
CC polymerase III complex (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- PTM: This protein undergoes a protein self splicing that involves
CC a post-translational excision of the intervening region (intein)
CC followed by peptide ligation.
CC -1- MISCELLANEOUS: The intein is a split intein capable of protein
CC trans-splicing. The N- and C-terminal halves of DnaE are encoded by
CC two separate genes located 745,226 bp apart in the genome and on
CC opposite DNA strands. The dnaE-n product consists of a N-extein
CC sequence followed by a 123-aa intein sequence (Ssp dnaE1), whereas
CC the dnaE-c product consists of a 36-aa intein sequence (Ssp dnaE2)
CC followed by a C-extein sequence. The N- and C-extein sequences
CC together reconstitute a complete DnaE sequence that is interrupted
CC by the intein sequences inside the beta- and tau-binding domains.
CC -1- SIMILARITY: Belongs to the DNA polymerase type-C family. DnaE
CC subfamily.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; BA000022; BAI8870.1; ALT_SEQ; Genomic DNA.
CC EMBL; BA000022; BAI17242.1; ALT_SEQ; Genomic DNA.
CC InterPro; IPR011708; DNA_pol3_alpha.
CC InterPro; IPR003587; Hedgehog_hint_N.
CC InterPro; IPR006141; Intein_S.
CC InterPro; IPR004365; OB_cRNA_NA_bd.
CC InterPro; IPR003141; Pectinase_PHP_N.
CC InterPro; IPR004013; PHP_C.
CC InterPro; IPR004805; PolC_alpha.
CC Pfam; PF07733; DNA_pol3_alpha; 1.
CC Pfam; PF02811; PHP1.1.
CC Pfam; PF01336; tRNA_anti_1.
CC SMART; SM00306; HintN; 1.
CC SMART; SM00481; POLIITAC; 1.
CC TIGRfams; TIGR01443; Intein_Cterm; 1.
CC TIGRfams; TIGR01445; Intein_Nterm; 1.
CC TIGRfams; TIGR00594; PolC; 1.
CC PROSITE; PS50818; INTEIN_C_TER; 1.
CC PROSITE; PS50817; INTEIN_N_TER; 1.
CC Autocatalytic cleavage; Complete proteome; DNA replication;
CC DNA-directed DNA polymerase; Nucleotidyltransferase; Protein splicing;
CC Transferase.
CC CHAIN 1 774
CC part.
FT

```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: May 4, 2006, 03:27:26 ; Search time 26.1111 Seconds
(Without alignments)
55.273 Million cell updates/sec

Title: US-10-799-326-78
Perfect score: 58
Sequence: 1 DGXVIRATXKHXFXT 15

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : PIR 80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	58	100.0	897 2	S76958 DNA-directed DNA p
2	54	93.1	877 2	AC2253 DNA polymerase III
3	46	79.3	613 2	T27528 hypothetical prote
4	46	79.3	879 2	A11878 replicative DNA he
5	46	79.3	1437 2	C75198 activator I, repli
6	45	77.6	872 2	S75781 replicative DNA he
7	42	72.4	434 2	H70989 probable GABA perm
8	42	72.4	550 2	S44980 flagellin - Shigel
9	42	72.4	561 2	JB0158 dwarf protein, OSD
10	42	72.4	893 2	S46442 nitrate reductase
11	42	72.4	1108 2	E71104 probable cell divi
12	41	70.7	647 2	A81406 DNA ligase (NAD) (
13	41	70.7	786 2	A69927 ribonucleoside-dip
14	41	70.7	1084 2	T12925 ribonucleotide red
15	41	70.7	1847 2	E64477 replication factor
16	40	69.0	114 2	S38226 hypothetical prote
17	40	69.0	152 2	E95018 conserved domain p
18	40	69.0	152 2	P97891 hypothetical prote
19	40	69.0	177 2	B69520 hypothetical prote
20	40	69.0	344 2	T14018 NADH2 dehydrogenas
21	40	69.0	811 2	S77577 endo-alpha-sialida
22	40	69.0	823 2	T02812 probable membrane
23	39	67.2	347 2	T11834 NADH2 dehydrogenas
24	39	67.2	379 1	B28940 cytochrome d ubiqn
25	39	67.2	379 1	A99725 cytochrome d ubiqn
26	39	67.2	379 2	B85576 cytochrome d ubiqn
27	39	67.2	379 2	AC0862 conserved hypothet
28	39	67.2	454 2	A85931 hypothetical prote
29	39	67.2	454 2	G65061

30	39	67.2	454 2	G91085 hypothetical prote
31	39	67.2	457 2	B82268 conserved hypothet
32	39	67.2	604 2	F87244 replicative DNA he
33	39	67.2	1042 2	H75112 mol/ydenum cofacto
34	39	67.2	1401 2	S11527 alpha-latrotoxin p
35	39	67.2	2150 2	T32497 hypothetical prote
36	38	65.5	236 2	AF2927 transcription regu
37	38	65.5	262 2	A98355 transcription regu
38	38	65.5	315 2	S44234 strN protein - Str
39	38	65.5	395 2	AE1801 D-3-phosphoglycer
40	38	65.5	395 2	AG1427 D-3-phosphoglycer
41	38	65.5	397 2	T35609 white protein I - S
42	38	65.5	438 2	E83938 allantoicase BH230
43	38	65.5	516 2	A75623 GGDEF family prote
44	38	65.5	519 2	AD2705 conserved hypothet
45	38	65.5	519 2	D97487 BH2174 hypothet
46	38	65.5	522 2	D96602 nucleolar protein
47	38	65.5	531 2	B83422 probable serine/th
48	38	65.5	702 2	T34313 hypothetical prote
49	38	65.5	775 1	VPXRF outer layer protei
50	38	65.5	908 1	UN0804 nitrate reductase
51	38	65.5	1024 2	T05234 hypothetical prote
52	38	65.5	1068 2	S73091 hypothetical prote
53	38	65.5	1345 2	A64430 DNA-directed RNA p
54	38	65.5	1352 2	G71051 probable ATP-depen
55	38	65.5	2122 2	B75009 ribonucleotide red
56	38	65.5	3084 1	MMMSA laminin alpha-1 ch
57	37.5	64.7	245 2	C87399 hypothetical prote
58	37.5	64.7	766 2	T47944 hypothetical prote
59	37.5	63.8	112 2	S57261 nitrate reductase
60	37	63.8	137 2	A75043 hypothetical prote
61	37	63.8	147 2	C47056 cnr regulatory pro
62	37	63.8	157 2	B71118 hypothetical prote
63	37	63.8	162 1	A40803 tropomyosin C, skelet
64	37	63.8	163 1	TPCHCS tropomyosin C, skelet
65	37	63.8	183 2	AE2498 hypothetical prote
66	37	63.8	222 2	E72261 conserved hypothet
67	37	63.8	231 2	F95267 hypothetical prote
68	37	63.8	254 2	T40159 casein kinase II,
69	37	63.8	275 2	G89961 hypothetical prote
70	37	63.8	296 2	F90067 hypothetical prote
71	37	63.8	317 2	JC7207 mpsA protein - Mag
72	37	63.8	325 2	G64055 aerobic respiratio
73	37	63.8	347 1	DNHUN2 NADH2 dehydrogenas
74	37	63.8	349 2	F95971 probable transcrip
75	37	63.8	369 2	AG0224 thermoregulated mo
76	37	63.8	378 2	JC5689 CIRCE binding repr
77	37	63.8	390 2	G72491 hypothetical prote
78	37	63.8	525 2	E66786 protein F10A5.13 l
79	37	63.8	559 2	H84859 hypothetical prote
80	37	63.8	572 2	T37128 hypothetical prote
81	37	63.8	616 2	C75588 conserved hypothet
82	37	63.8	628 2	S27586 ab1829 protein - L
83	37	63.8	658 1	S73805 DNA ligase (NAD) (
84	37	63.8	689 2	T09007 ABC-transporter ho
85	37	63.8	739 1	D87437 TonB-dependent rec
86	37	63.8	739 1	A70204 CT456 hypothet
87	37	63.8	755 2	H86561 hypothetical prote
88	37	63.8	755 2	B72061 hypothetical prote
89	37	63.8	790 2	A82200 cation transport A
90	37	63.8	795 2	D64343 hypothetical prote
91	37	63.8	859 2	S70584 nitrate reductase
92	37	63.8	867 1	JQ1525 nitrate reductase
93	37	63.8	868 1	JC4283 nitrate reductase
94	37	63.8	873 1	JH0182 nitrate reductase
95	37	63.8	894 2	S52857 nitrate reductase
96	37	63.8	905 1	UN0803 nitrate reductase
97	37	63.8	982 1	S16292 nitrate reductase
98	37	63.8	1433 1	GANVUB M polypeptide prec
99	37	63.8	1479 2	T42710 mannose receptor,
100	37	63.8	1524 2	T30518 carbamoyl-phosphat

ALIGNMENTS

```
RESULT 1
576958
DNA-directed DNA polymerase (EC 2.7.7.7) III alpha chain split extein-N containing precu
C/Species: Synecchococcus sp.
A/Variety: PCC 6803
C/Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 12-Jul-2004
C/Accession: S76958
R/Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
o, K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3: 109-116, 1996
A/Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecchocystis
s.
A/Reference number: S74322; MUID:97061201; PMID:8905231
A/Accession: S76958
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-897 <KAN>
A/Cross-references: UNIPARC:UPI0000164C87; EMBL:D90917; GB:AB001339; NID:g163836; PIDN:
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
R:Wu, H.; Hu, Z.; Lu, X.Q.
Proc. Natl. Acad. Sci. U.S.A. 95, 9226-9231, 1998
A/Title: Protein trans-splicing by a split intem encoded in a split DnaE gene of Synecch
A/Reference number: A59016; MUID:96356139; PMID:9689062
A/Contents: annotation
C/Comment: For the other split extein precursor see PIR:S75328. For the mature spliced f
C/Genetics:
A/Gene: dnaE-n
C/Keywords: DNA biosynthesis; DNA replication; nucleotidyltransferase; protein splicing
P:1-776/Domain: DNA-directed DNA polymerase III alpha chain split extein-N #status predi
P:775-897/Domain: split intein-N #status predicted <NTN>
Query Match 100.0%; Score 58; DB 2; Length 897;
Best Local Similarity 66.7%; Pred. No. 0.08;
Matches 10; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
QY 1 DGXVIRATXXHXFXT 15
||:|||||:|:|
Db 836 DGSVIRATSDHRLF 850

RESULT 2
AC2253
DNA polymerase III alpha chain [imported] - Nostoc sp. (strain PCC 7120)
C/Species: Nostoc sp. PCC 7120
A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C/Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 12-Jul-2004
C/Accession: AC2253
R/Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi
Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S
DNA Res. 8: 205-213, 2001
A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A/Reference number: AB1807; MUID:21595285; PMID:11759840
A/Accession: AC2253
A/Status: Preliminary
A/Molecule type: DNA
A/Residues: 1-877 <KUR>
A/Cross-references: UNIPROT:Q8YR72; UNIPARC:UPI000000CB838; GB:BA000019; PIDN:BAB75277.1;
A/Experimental source: strain PCC 7120
C/Genetics:
A/Gene: all3578
Query Match 93.1%; Score 54; DB 2; Length 877;
Best Local Similarity 53.3%; Pred. No. 0.41;
Matches 8; Conservative 7; Mismatches 0; Indels 0; Gaps 0;
QY 1 DGXVIRATXXHXFXT 15
||:|||||:|:|
Db 837 DGSIRATKHKHFWT 851
```

```
RESULT 3
T27528
hypothetical protein ZC374.2 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C/Accession: T27528
R/White, S.
submitted to the EMBL Data Library, May 1996
A/Reference number: Z20383
A/Accession: T27528
A/Status: Preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-613 <WIL>
A/Cross-references: UNIPROT:Q23266; UNIPARC:UPI000007879B; EMBL:Z72518; PIDN:CAA96700.1
A/Experimental source: clone ZC374
C/Genetics:
A/Gene: CESP:ZC374.2
A/Map position: X
A/Intons: 48/1; 90/3; 114/3; 219/1; 241/3; 287/1; 326/1; 359/2; 389/1; 434/3; 490/1; 5
Query Match 79.3%; Score 46; DB 2; Length 613;
Best Local Similarity 53.3%; Pred. No. 7.5;
Matches 8; Conservative 6; Mismatches 1; Indels 0; Gaps 0;
QY 1 DGXVIRATXXHXFXT 15
||:|||||:|:|
Db 474 DGSVIRATSVRFXFT 488

RESULT 4
A11878
replicative DNA helicase [imported] - Nostoc sp. (strain PCC 7120)
C/Species: Nostoc sp. PCC 7120
A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C/Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C/Accession: A11878
R/Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi
Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S
DNA Res. 8: 205-213, 2001
A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium An
A/Reference number: AB1807; MUID:21595285; PMID:11759840
A/Accession: A11878
A/Status: Preliminary
A/Molecule type: DNA
A/Residues: 1-879 <KUR>
A/Cross-references: UNIPROT:Q8Y2A1; UNIPARC:UPI0000129550; GB:BA000019; PIDN:BAB72536.1
A/Experimental source: strain PCC 7120
C/Genetics:
A/Gene: dnaB
Query Match 79.3%; Score 46; DB 2; Length 879;
Best Local Similarity 57.1%; Pred. No. 11;
Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
QY 2 GXYIRATXXHXFXT 15
||:|||||:|:|
Db 452 GCKIRATNKHFWT 465

RESULT 5
C75198
activator 1, replication factor c, small chain PAB0068 - Pyrococcus abyssi (strain Orsa
C/Species: Pyrococcus abyssi
C/Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
C/Accession: C75198
R/anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A/Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome str
A/Reference number: A75001
A/Accession: C75198
A/Status: Preliminary
A/Molecule type: DNA
A/Residues: 1-1437 <KAN>
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: May 4, 2006, 03:34:13 ; Search time 19.4444 Seconds
(without alignments)
35.705 Million cell updates/sec

Title: US-10-799-326-78

Perfect score: 58

Sequence: 1 DGXVIRATXXHFXFT 15

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 235405 seqs, 46284737 residues

Total number of hits satisfying chosen parameters: 235405

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Published Applications AA New:*

1: /SID55/ptodata/1/pubpaa/US06_NEW_PUB.pep:1*
2: /SID55/ptodata/1/pubpaa/US07_NEW_PUB.pep:1*
3: /SID55/ptodata/1/pubpaa/US08_NEW_PUB.pep:1*
4: /SID55/ptodata/1/pubpaa/US09_NEW_PUB.pep:1*
5: /SID55/ptodata/1/pubpaa/US10_NEW_PUB.pep:1*
6: /SID55/ptodata/1/pubpaa/US11_NEW_PUB.pep:1*
7: /SID55/ptodata/1/pubpaa/US12_NEW_PUB.pep:1*
8: /SID55/ptodata/1/pubpaa/US13_NEW_PUB.pep:1*
9: /SID55/ptodata/1/pubpaa/US14_NEW_PUB.pep:1*
10: /SID55/ptodata/1/pubpaa/US15_NEW_PUB.pep:1*
11: /SID55/ptodata/1/pubpaa/US16_NEW_PUB.pep:1*
12: /SID55/ptodata/1/pubpaa/US17_NEW_PUB.pep:1*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	58	100.0	124	11	US-11-069-642-10
2	46	79.3	610	11	US-11-069-642-22
3	45	77.6	155	11	US-11-069-642-97
4	45	77.6	198	11	US-11-069-642-25
5	45	77.6	431	11	US-11-069-642-1
6	45	77.6	437	11	US-11-069-642-92
7	45	77.6	438	11	US-11-069-642-47
8	45	77.6	438	11	US-11-069-642-49
9	45	77.6	438	11	US-11-069-642-51
10	45	77.6	438	11	US-11-069-642-53
11	45	77.6	438	11	US-11-069-642-55
12	45	77.6	438	11	US-11-069-642-57
13	45	77.6	438	11	US-11-069-642-59
14	45	77.6	438	11	US-11-069-642-61
15	45	77.6	438	11	US-11-069-642-63
16	45	77.6	949	11	US-11-077-550-68
17	45	77.6	1420	11	US-11-077-550-110
18	42	72.4	434	11	US-11-087-099-6812
19	42	72.4	434	11	US-11-188-298-17245
20	42	72.4	550	11	US-11-052-554A-200
21	41	70.7	325	11	US-11-096-568A-25271

22	41	70.7	358	11	US-11-096-568A-25270	Sequence 25270, A
23	41	70.7	387	11	US-11-096-568A-25269	Sequence 25269, A
24	40	69.0	681	11	US-11-079-463-6507	Sequence 6507, Ap
25	39	67.2	147	11	US-11-069-642-11	Sequence 11, Appl
26	39	67.2	171	9	US-10-506-454-254	Sequence 254, App
27	39	67.2	261	11	US-11-098-686-10986	Sequence 10986, A
28	39	67.2	1008	11	US-11-079-463-8074	Sequence 8074, Ap
29	39	67.2	1119	11	US-11-079-463-9958	Sequence 9958, Ap
30	38	65.5	395	11	US-11-045-004-99	Sequence 99, Appl
31	38	65.5	522	11	US-11-087-099-284	Sequence 284, App
32	38	65.5	1257	9	US-10-506-454-679	Sequence 679, App
33	37	63.8	191	9	US-10-467-657-818	Sequence 818, App
34	37	63.8	233	11	US-11-096-568A-24646	Sequence 24646, A
35	37	63.8	237	11	US-11-096-568A-24645	Sequence 24645, A
36	37	63.8	254	11	US-11-188-298-4554	Sequence 4554, Ap
37	37	63.8	284	11	US-11-188-298-19909	Sequence 19909, A
38	37	63.8	454	11	US-11-059-867-3	Sequence 3, Appl1
39	37	63.8	555	11	US-11-188-298-6221	Sequence 6221, Ap
40	36	62.1	132	9	US-10-821-234-1218	Sequence 1218, Ap
41	36	62.1	144	11	US-11-072-512-3457	Sequence 3457, Ap
42	36	62.1	168	11	US-11-188-298-6922	Sequence 6922, Ap
43	36	62.1	200	9	US-10-506-454-1468	Sequence 1468, App
44	36	62.1	241	11	US-11-082-389-20	Sequence 20, Appl
45	36	62.1	273	11	US-11-188-298-19626	Sequence 19626, A
46	36	62.1	282	11	US-11-096-568A-5326	Sequence 5326, Ap
47	36	62.1	309	11	US-11-096-568A-5325	Sequence 5325, Ap
48	36	62.1	335	11	US-11-096-568A-5324	Sequence 5324, Ap
49	36	62.1	373	11	US-11-188-298-7205	Sequence 7205, Ap
50	36	62.1	379	11	US-11-188-298-6617	Sequence 6617, Ap
51	36	62.1	379	11	US-10-878-556A-116	Sequence 116, App
52	36	62.1	465	9	US-10-194-487-316	Sequence 316, App
53	36	62.1	466	9	US-10-195-883-316	Sequence 316, App
54	36	62.1	466	9	US-10-195-883-316	Sequence 316, App
55	36	62.1	466	9	US-10-195-883-316	Sequence 316, App
56	36	62.1	466	9	US-10-195-883-316	Sequence 316, App
57	36	62.1	466	9	US-10-195-883-316	Sequence 316, App
58	36	62.1	466	11	US-11-102-240-104	Sequence 104, App
59	36	62.1	466	11	US-11-173-672-1	Sequence 1, Appl1
60	36	62.1	466	11	US-11-103-195-104	Sequence 104, App
61	36	62.1	534	9	US-10-646-283-2	Sequence 2, Appl1
62	36	62.1	615	11	US-11-136-244-13	Sequence 13, Appl
63	36	62.1	650	9	US-10-467-657-1948	Sequence 1948, App
64	36	62.1	1114	9	US-10-469-469-277	Sequence 277, App
65	35	60.3	40	11	US-11-053-185-43	Sequence 43, Appl
66	35	60.3	94	11	US-11-093-808-2	Sequence 2, Appl1
67	35	60.3	140	9	US-10-467-657-2486	Sequence 2486, Ap
68	35	60.3	189	11	US-11-087-057-1515	Sequence 1515, A
69	35	60.3	215	9	US-10-793-626-828	Sequence 828, App
70	35	60.3	305	9	US-10-506-454-55	Sequence 55, Appl
71	35	60.3	331	11	US-11-096-568A-22525	Sequence 22525, A
72	35	60.3	334	11	US-11-096-568A-10349	Sequence 10349, A
73	35	60.3	336	11	US-11-096-568A-22524	Sequence 22524, A
74	35	60.3	340	11	US-11-096-568A-24096	Sequence 24096, A
75	35	60.3	351	11	US-11-096-568A-22523	Sequence 22523, A
76	35	60.3	351	11	US-11-096-568A-24095	Sequence 24095, A
77	35	60.3	361	11	US-11-096-568A-24094	Sequence 24094, A
78	35	60.3	371	11	US-11-096-568A-10348	Sequence 10348, A
79	35	60.3	398	9	US-10-966-628-4	Sequence 4, Appl1
80	35	60.3	398	11	US-11-223-492-6	Sequence 6, Appl1
81	35	60.3	412	11	US-11-093-808-1	Sequence 1, Appl1
82	35	60.3	412	11	US-11-093-808-8	Sequence 8, Appl1
83	35	60.3	412	11	US-11-093-808-9	Sequence 9, Appl1
84	35	60.3	412	11	US-11-093-808-10	Sequence 10, Appl1
85	35	60.3	412	11	US-11-093-808-11	Sequence 11, Appl1
86	35	60.3	412	11	US-11-093-808-12	Sequence 12, Appl1
87	35	60.3	412	11	US-11-093-808-13	Sequence 13, Appl1
88	35	60.3	412	11	US-11-093-808-14	Sequence 14, Appl1
89	35	60.3	431	11	US-11-079-463-7041	Sequence 7041, Ap
90	35	60.3	432	9	US-10-878-556A-127	Sequence 127, App
91	35	60.3	566	11	US-11-072-512-2547	Sequence 2547, Ap
92	35	60.3	652	9	US-10-506-454-1255	Sequence 1255, Ap
93	35	60.3	753	11	US-11-072-512-2512	Sequence 2512, Ap
94	35	60.3	864	11	US-11-079-463-5554	Sequence 5554, Ap

95 35 60.3 1249 9 US-10-506-454-486 Sequence 486, App
96 35 60.3 1813 9 US-10-495-083-10 Sequence 10, App
97 34 58.6 23 11 US-11-069-642-44 Sequence 44, App
98 34 58.6 23 11 US-11-069-642-85 Sequence 85, App
99 34 58.6 69 11 US-11-237-597-44 Sequence 44, App
100 34 58.6 69 11 US-11-096-568A-3014 Sequence 3014, App

ALIGNMENTS

RESULT 1

US-11-069-642-10
; Sequence 10, Application US/11069642
; Publication No. US20050260626A1
; GENERAL INFORMATION:
; APPLICANT: LORENS, JAMES B.
; APPLICANT: KINSELLA, TODD M.
; APPLICANT: BENNETT, MARK K.
; TITLE OF INVENTION: IN VIVO PRODUCTION OF CYCLIC PEPTIDES FOR
; FILE REFERENCE: RIGL-022CIP3
; CURRENT APPLICATION NUMBER: US/11/069,642
; CURRENT FILING DATE: 2005-02-28
; PRIOR APPLICATION NUMBER: 10/232,758
; PRIOR FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: 10/422,536
; PRIOR FILING DATE: 2003-04-23
; PRIOR APPLICATION NUMBER: 09/800,770
; PRIOR FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: 60/187,130
; PRIOR FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 168
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Synechocystis PCC6803
US-11-069-642-10

Query Match 100.0%; Score 58; DB 11; Length 124;
Best Local Similarity 66.7%; Pred. No. 0.0071;
Matches 10; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGXVIRATXXHFXFT 15

Db 63 DGSVIRATSDHRLFT 77

RESULT 2

US-11-069-642-22
; Sequence 22, Application US/11069642
; Publication No. US20050260626A1
; GENERAL INFORMATION:
; APPLICANT: LORENS, JAMES B.
; APPLICANT: PRAY, TODD R.
; APPLICANT: KINSELLA, TODD M.
; APPLICANT: BENNETT, MARK K.
; TITLE OF INVENTION: IN VIVO PRODUCTION OF CYCLIC PEPTIDES FOR
; FILE REFERENCE: RIGL-022CIP3
; CURRENT APPLICATION NUMBER: US/11/069,642
; CURRENT FILING DATE: 2005-02-28
; PRIOR APPLICATION NUMBER: 10/232,758
; PRIOR FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: 10/422,536
; PRIOR FILING DATE: 2003-04-23
; PRIOR APPLICATION NUMBER: 09/800,770
; PRIOR FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: 60/187,130
; PRIOR FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 168

; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 22
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Pyrococcus abyssi
US-11-069-642-22

Query Match 79.3%; Score 46; DB 11; Length 610;
Best Local Similarity 50.0%; Pred. No. 6.3;
Matches 7; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 2 GXVIRATXXHFXFT 15

Db 66 GRITKATGDHRLFT 79

RESULT 3

US-11-069-642-97
; Sequence 97, Application US/11069642
; Publication No. US20050260626A1
; GENERAL INFORMATION:
; APPLICANT: LORENS, JAMES B.
; APPLICANT: KINSELLA, TODD M.
; APPLICANT: BENNETT, MARK K.
; TITLE OF INVENTION: IN VIVO PRODUCTION OF CYCLIC PEPTIDES FOR
; FILE REFERENCE: RIGL-022CIP3
; CURRENT APPLICATION NUMBER: US/11/069,642
; CURRENT FILING DATE: 2005-02-28
; PRIOR APPLICATION NUMBER: 10/232,758
; PRIOR FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: 10/422,536
; PRIOR FILING DATE: 2003-04-23
; PRIOR APPLICATION NUMBER: 09/800,770
; PRIOR FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: 60/187,130
; PRIOR FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 168
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 97
; LENGTH: 155
; TYPE: PRT
; ORGANISM: Synechocystis PCC6803
US-11-069-642-97

Query Match 77.6%; Score 45; DB 11; Length 155;
Best Local Similarity 50.0%; Pred. No. 1.9;
Matches 7; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 2 GXVIRATXXHFXFT 15

Db 64 GRITKATANHRLFT 77

RESULT 4

US-11-069-642-25
; Sequence 25, Application US/11069642
; Publication No. US20050260626A1
; GENERAL INFORMATION:
; APPLICANT: LORENS, JAMES B.
; APPLICANT: PRAY, TODD R.
; APPLICANT: KINSELLA, TODD M.
; APPLICANT: BENNETT, MARK K.
; TITLE OF INVENTION: IN VIVO PRODUCTION OF CYCLIC PEPTIDES FOR
; FILE REFERENCE: RIGL-022CIP3
; CURRENT APPLICATION NUMBER: US/11/069,642
; CURRENT FILING DATE: 2005-02-28
; PRIOR APPLICATION NUMBER: 10/232,758
; PRIOR FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: 10/422,536
; PRIOR FILING DATE: 2003-04-23

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: May 4, 2006, 03:33:07 ; Search time 133.333 Seconds
(without alignments)
47.006 Million cell updates/sec

Title: US-10-799-326-78

Perfect score: 58

Sequence: 1 DGXVIRATXXHXFT 15

Scoring table: BLOSUM62DX

Searched: Gapex 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Published Applications AA Main:*

- 1: /cgn2_6/prodata/1/pubppa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/prodata/1/pubppa/US08_PUBCOMB.pep:*
- 3: /cgn2_6/prodata/1/pubppa/US09_PUBCOMB.pep:*
- 4: /cgn2_6/prodata/1/pubppa/US10A_PUBCOMB.pep:*
- 5: /cgn2_6/prodata/1/pubppa/US10B_PUBCOMB.pep:*
- 6: /cgn2_6/prodata/1/pubppa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	58	100.0	15	4	US-10-799-326-74
2	58	100.0	15	4	US-10-799-326-76
3	58	100.0	15	4	US-10-799-326-78
4	58	100.0	78	4	US-10-799-326-71
5	58	100.0	123	4	US-10-356-088-1
6	58	100.0	123	4	US-10-356-088-23
7	58	100.0	123	4	US-10-353-454-1
8	58	100.0	123	4	US-10-353-454-23
9	58	100.0	123	4	US-10-799-326-1
10	58	100.0	123	4	US-10-799-326-23
11	58	100.0	124	3	US-09-800-770-10
12	58	100.0	124	6	US-11-011-880-10
13	58	100.0	124	6	US-11-011-880-10
14	58	100.0	320	4	US-10-356-088-39
15	58	100.0	320	4	US-10-356-088-39
16	58	100.0	326	4	US-10-356-088-38
17	58	100.0	326	4	US-10-799-326-38
18	58	100.0	416	5	US-10-833-951-25
19	58	100.0	461	5	US-10-833-951-4
20	46	79.3	610	3	US-09-800-770-22
21	46	79.3	610	4	US-10-422-536-22
22	46	79.3	610	6	US-11-011-880-22
23	46	79.3	1437	4	US-10-369-493-1237
24	45	77.6	14	5	US-10-833-951-98
25	45	77.6	95	4	US-10-460-594-78
26	45	77.6	155	3	US-09-800-770-54
27	45	77.6	155	4	US-10-422-536-97

28	45	77.6	155	6	US-11-011-880-54	Sequence 54, Appl
29	45	77.6	198	3	US-09-800-770-25	Sequence 25, Appl
30	45	77.6	198	4	US-10-422-536-25	Sequence 25, Appl
31	45	77.6	198	6	US-11-011-880-25	Sequence 25, Appl
32	45	77.6	199	3	US-09-800-770-28	Sequence 28, Appl
33	45	77.6	199	6	US-11-011-880-28	Sequence 28, Appl
34	45	77.6	431	3	US-09-800-770-1	Sequence 1, Appl
35	45	77.6	431	4	US-10-422-536-1	Sequence 1, Appl
36	45	77.6	431	6	US-11-011-880-1	Sequence 1, Appl
37	45	77.6	434	3	US-09-800-770-48	Sequence 48, Appl
38	45	77.6	434	6	US-11-011-880-48	Sequence 48, Appl
39	45	77.6	437	3	US-09-800-770-44	Sequence 44, Appl
40	45	77.6	437	4	US-10-422-536-92	Sequence 92, Appl
41	45	77.6	437	6	US-11-011-880-49	Sequence 49, Appl
42	45	77.6	438	3	US-09-800-770-30	Sequence 30, Appl
43	45	77.6	438	6	US-10-422-536-30	Sequence 30, Appl
44	45	77.6	438	3	US-09-800-770-32	Sequence 32, Appl
45	45	77.6	438	6	US-11-011-880-32	Sequence 32, Appl
46	45	77.6	438	3	US-09-800-770-34	Sequence 34, Appl
47	45	77.6	438	6	US-10-422-536-47	Sequence 47, Appl
48	45	77.6	438	3	US-09-800-770-40	Sequence 40, Appl
49	45	77.6	438	6	US-11-011-880-40	Sequence 40, Appl
50	45	77.6	438	3	US-09-800-770-44	Sequence 44, Appl
51	45	77.6	438	6	US-10-422-536-55	Sequence 55, Appl
52	45	77.6	438	3	US-09-800-770-57	Sequence 57, Appl
53	45	77.6	438	6	US-11-011-880-57	Sequence 57, Appl
54	45	77.6	438	3	US-10-422-536-59	Sequence 59, Appl
55	45	77.6	438	6	US-11-011-880-59	Sequence 59, Appl
56	45	77.6	438	3	US-09-800-770-63	Sequence 63, Appl
57	45	77.6	438	6	US-10-422-536-63	Sequence 63, Appl
58	45	77.6	438	3	US-09-800-770-34	Sequence 34, Appl
59	45	77.6	438	6	US-11-011-880-34	Sequence 34, Appl
60	45	77.6	438	3	US-10-422-536-36	Sequence 36, Appl
61	45	77.6	438	6	US-11-011-880-36	Sequence 36, Appl
62	45	77.6	438	3	US-09-800-770-46	Sequence 46, Appl
63	45	77.6	438	6	US-10-422-536-46	Sequence 46, Appl
64	45	77.6	438	3	US-09-800-770-42	Sequence 42, Appl
65	45	77.6	438	6	US-11-011-880-42	Sequence 42, Appl
66	45	77.6	438	3	US-09-800-770-44	Sequence 44, Appl
67	45	77.6	438	6	US-10-422-536-44	Sequence 44, Appl
68	45	77.6	438	3	US-09-800-770-46	Sequence 46, Appl
69	45	77.6	438	6	US-11-011-880-46	Sequence 46, Appl
70	45	77.6	1420	4	US-10-241-596-68	Sequence 68, Appl
71	43	74.1	131	4	US-10-767-701-6067	Sequence 6067, A
72	42	72.4	106	4	US-10-460-594-76	Sequence 76, Appl
73	42	72.4	434	4	US-10-282-122A-62664	Sequence 62664, A
74	42	72.4	434	4	US-10-282-122A-64955	Sequence 64955, A
75	42	72.4	515	4	US-10-767-701-46978	Sequence 46978, A
76	42	72.4	550	4	US-10-125-692-22	Sequence 22, Appl
77	42	72.4	550	5	US-10-991-347-22	Sequence 22, Appl
78	42	72.4	561	4	US-10-259-194A-334	Sequence 334, Appl
79	42	72.4	561	4	US-10-395-463-20	Sequence 20, Appl
80	42	72.4	561	5	US-10-739-930-10033	Sequence 10033, A
81	42	72.4	1108	4	US-10-369-493-1247	Sequence 1247, Ap
82	42	72.4	1108	4	US-10-369-493-20314	Sequence 20314, A
83	41	70.7	14	5	US-10-833-951-129	Sequence 129, Appl
84	41	70.7	107	4	US-10-460-594-85	Sequence 85, Appl
85	41	70.7	153	4	US-10-437-963-16765	Sequence 16765, A
86	41	70.7	447	4	US-10-425-115-20648	Sequence 20648, A
87	41	70.7	482	4	US-10-425-115-65468	Sequence 65468, A
88	41	70.7	618	4	US-10-282-122A-49670	Sequence 49670, A
89	41	70.7	647	4	US-10-282-122A-54370	Sequence 54370, A
90	41	70.7	1847	4	US-10-369-493-1075	Sequence 1075, Ap
91	40	69.0	70	4	US-10-437-963-181933	Sequence 121933, A
92	40	69.0	79	4	US-10-424-559-247044	Sequence 247044, A
93	40	69.0	152	4	US-10-474-776-2231	Sequence 2231, Appl
94	40	69.0	152	5	US-10-472-928-126	Sequence 126, Appl
95	40	69.0	181	4	US-10-437-963-115203	Sequence 115203, A
96	40	69.0	184	5	US-10-450-763-49203	Sequence 49203, A
97	40	69.0	209	5	US-10-617-320-3477	Sequence 3477, Ap
98	40	69.0	2404	6	US-11-097-143-34419	Sequence 34419, A
99	39	67.2	78	4	US-10-424-559-154585	Sequence 154585, A
100	39	67.2	108	4	US-10-425-115-277978	Sequence 277978, A

ALIGNMENTS

RESULT 1
US-10-799-326-74

Sequence 74, Application US/10799326
Publication No. US20040172688A1
GENERAL INFORMATION:
APPLICANT: E.I. DuPont de Nemours, & Company
APPLICANT: Yadav, Narendra S.
TITLE OF INVENTION: Intein-Mediated Protein Splicing
FILE REFERENCE: CL1806 US CIP
CURRENT FILING DATE: 2004-03-12
PRIORITY FILING DATE: 2004-03-12
PRIORITY FILING DATE: 2002-02-04
NUMBER OF SEQ ID NOS: 78
SOFTWARE: PatentIn version 3.2
SEQ ID NO 74
LENGTH: 15
TYPE: PRT
ORGANISM: *Synechocystis* sp. PCC6803
US-10-799-326-74

Query Match 100.0%; Score 58; DB 4; Length 15;
Best Local Similarity 66.7%; Pred. No. 0.0059;
Matches 10; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGXVIRATXXHXFXT 15
DB 1 DGSVIRATSDHRRFLT 15

RESULT 2
US-10-799-326-76

Sequence 76, Application US/10799326
Publication No. US20040172688A1
GENERAL INFORMATION:
APPLICANT: E.I. DuPont de Nemours, & Company
APPLICANT: Yadav, Narendra S.
APPLICANT: Yang, Jianjun Gene
TITLE OF INVENTION: Intein-Mediated Protein Splicing
FILE REFERENCE: CL1806 US CIP
CURRENT FILING DATE: 2004-03-12
PRIORITY FILING DATE: 2004-03-12
PRIORITY FILING DATE: 2002-02-04
NUMBER OF SEQ ID NOS: 78
SOFTWARE: PatentIn version 3.2
SEQ ID NO 76
LENGTH: 15
TYPE: PRT
ORGANISM: *Trichodesmium erythraeum*
US-10-799-326-76

Query Match 100.0%; Score 58; DB 4; Length 15;
Best Local Similarity 66.7%; Pred. No. 0.0059;
Matches 10; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGXVIRATXXHXFXT 15
DB 1 DGVIRATPBHKEFWT 15

RESULT 3
US-10-799-326-78

Sequence 78, Application US/10799326
Publication No. US20040172688A1
GENERAL INFORMATION:
APPLICANT: E.I. DuPont de Nemours, & Company

APPLICANT: Yadav, Narendra S.

APPLICANT: Yang, Jianjun Gene

TITLE OF INVENTION: Intein-Mediated Protein Splicing

FILE REFERENCE: CL1806 US CIP

CURRENT FILING DATE: 2004-03-12

PRIORITY FILING DATE: 2004-03-12

PRIORITY FILING DATE: 2002-02-04

NUMBER OF SEQ ID NOS: 78

SOFTWARE: PatentIn version 3.2

SEQ ID NO 78

LENGTH: 15

TYPE: PRT

ORGANISM: artificial

FEATURE:

OTHER INFORMATION: compiled sequence

FEATURE:

NAME/KEY: misc_feature

LOCATION: (3)..(13)

OTHER INFORMATION: Xaa can be any naturally occurring amino acid

FEATURE:

NAME/KEY: misc_feature

LOCATION: (12)..(12)

OTHER INFORMATION: Xaa can be any naturally occurring amino acid

FEATURE:

NAME/KEY: misc_feature

LOCATION: (14)..(14)

OTHER INFORMATION: Xaa can be any naturally occurring amino acid

US-10-799-326-78

Query Match 100.0%; Score 58; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0059;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGXVIRATXXHXFXT 15
DB 1 DGXVIRATXXHXFXT 15

RESULT 4
US-10-799-326-71

Sequence 71, Application US/10799326
Publication No. US20040172688A1
GENERAL INFORMATION:
APPLICANT: E.I. DuPont de Nemours, & Company
APPLICANT: Yadav, Narendra S.
APPLICANT: Yang, Jianjun Gene
TITLE OF INVENTION: Intein-Mediated Protein Splicing
FILE REFERENCE: CL1806 US CIP
CURRENT FILING DATE: 2004-03-12
PRIORITY FILING DATE: 2004-03-12
PRIORITY FILING DATE: 2002-02-04
NUMBER OF SEQ ID NOS: 78
SOFTWARE: PatentIn version 3.2
SEQ ID NO 71
LENGTH: 78
TYPE: PRT
ORGANISM: *Synechocystis* sp. PCC6803
US-10-799-326-71

Query Match 100.0%; Score 58; DB 4; Length 78;
Best Local Similarity 66.7%; Pred. No. 0.043;
Matches 10; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGXVIRATXXHXFXT 15
DB 62 DGSVIRATSDHRRFLT 76

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: May 4, 2006, 03:17:19 ; Search time 130.667 Seconds
(without alignments)

64.793 Million cell updates/sec

Title: US-10-799-326-77

Perfect score: 45

Sequence: 1 CLXXXTEIXIVE 12

Scoring table: BLOSUM62DX

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

1: uniprot_sprot:*

2: uniprot_tr embl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	100.0	503	2	Q6XUH7_9NOST
2	45	100.0	712	2	Q6XUH8_ALPHA
3	45	100.0	1355	1	DP03A_SYNY3
4	44	97.8	137	2	Q6XUH3_9CYAN
5	44	97.8	870	2	Q5N0T2_SYNP6
6	44	97.8	877	2	Q6YR72_ANASP
7	40	88.9	817	2	Q66047_DROME
8	40	88.9	817	2	Q9V461_DROME
9	39	86.7	279	2	Q5CUN0_CRYPV
10	39	86.7	399	1	GR59E_DROME
11	39	86.7	426	2	Q6ZH07_YERPE
12	39	86.7	426	2	Q66DA7_YERPS
13	39	86.7	500	2	Q5TNU0_ANOGA
14	39	86.7	813	1	Q7Q0Q1_ANOGA
15	38	84.4	71	1	Y16X_BPT4
16	38	84.4	122	2	Q6W5B8_SYNVU
17	38	84.4	510	2	Q6FN68_CANGA
18	38	84.4	510	2	Q6VUP1_CANGA
19	38	84.4	872	2	Q6BDH1_SYNEL
20	37	82.2	142	2	Q6LTR8_PHOPR
21	37	82.2	184	2	Q9D6D4_MOUSE
22	37	82.2	184	2	Q510V9_MOUSE
23	37	82.2	203	2	Q98G93_RHILO
24	37	82.2	429	2	Q5KES0_CRYNE
25	37	82.2	429	2	Q5SRP0_CRYNE
26	37	82.2	445	2	Q5FHP6_EHRRG
27	37	82.2	544	1	PWE3_LYCES
28	37	82.2	585	2	Q5HHB0_EHRRW
29	36	80.0	65	2	Q89Y12_BACTN
30	36	80.0	73	2	Q6D2G8_ERWCT
31	36	80.0	112	2	Q08497_BARCL

32	36	80.0	112	2	Q5PSB1_9RHIZ	Q5PSB1_uncultured
33	36	80.0	193	1	HAMI_TROW8	Q83hf7_tropherya
34	36	80.0	193	1	HAMI_TROWT	Q83f12_tropherya
35	36	80.0	308	2	Q8CIE6_MOUSE	Q8CIE6_mus muscu
36	36	80.0	328	2	Q87617_PSEAE	Q87617_pseudomonas
37	36	80.0	328	2	Q5GRCO_ALCXX	Q5GRCO_alcaligenes
38	36	80.0	369	2	Q98R45_MYCPU	Q98R45_mycoplasm
39	36	80.0	426	2	Q5Q027_IDILO	Q5Q027_idiomarina
40	36	80.0	441	2	Q659R7_SACOF	Q659R7_saccharum o
41	36	80.0	446	1	HNF4B_XENLA	Q926n4_xenopus lae
42	36	80.0	526	1	G6PI_CHLPN	Q5XGTS_xenopus lae
43	36	80.0	682	2	Q5XGT5_XENLA	Q5XGTS_xenopus lae
44	36	80.0	692	2	Q8B252_MOUSE	Q8B252_mus muscu
45	36	80.0	800	2	Q7PX12_ANOGA	Q7PX12_anopheles g
46	36	80.0	809	2	Q4RFRI_TETNG	Q4RFRI_tetradodon n
47	36	80.0	815	2	Q6YON8_ANASP	Q6YON8_anabaena sp
48	36	80.0	832	2	Q7XCNO_ORYSA	Q7XCNO_oryza sativ
49	36	80.0	832	2	Q9PRK1_ORYSA	Q9PRK1_oryza sativ
50	36	80.0	850	2	Q4PIV4_USTMA	Q4PIV4_usstlago ma
51	36	80.0	1005	2	Q6CMB8_KULUA	Q6CMB8_kluyveromyc
52	36	80.0	1013	2	Q6CT40_KULUA	Q6CT40_kluyveromyc
53	36	80.0	1145	2	Q54XA7_DICDI	Q54XA7_dicystostell
54	36	80.0	1495	2	Q5A819_CANAL	Q5A819_candida alb
55	36	80.0	1809	2	Q15892_TETTH	Q15892_tetrahymena
56	36	80.0	5158	2	Q5D6D6_COCHC	Q5D6D6_cochliobol
57	36	80.0	6048	2	Q3JH87_STRAW	Q3JH87_streptomyce
58	36	80.0	7746	2	Q82QT4_STRAW	Q82QT4_streptomyce
59	35	77.8	95	2	Q9REZ2_9RHIZ	Q9REZ2_uncultured
60	35	77.8	104	1	YGV9_YEAST	P53109_saccharomyc
61	35	77.8	112	2	Q5PSJ1_9RHIZ	Q5PSJ1_uncultured
62	35	77.8	112	2	Q6T5G3_9RHIZ	Q6T5G3_bartonella
63	35	77.8	112	2	Q79BA0_9RHIZ	Q79BA0_bartonella
64	35	77.8	112	2	Q79BA1_9RHIZ	Q79BA1_bartonella
65	35	77.8	112	2	Q8KXW6_9RHIZ	Q8KXW6_bartonella
66	35	77.8	112	2	Q8KXW7_9RHIZ	Q8KXW7_bartonella
67	35	77.8	112	2	Q93M25_9RHIZ	Q93M25_bartonella
68	35	77.8	112	2	Q93M26_9RHIZ	Q93M26_bartonella
69	35	77.8	112	2	Q93M28_9RHIZ	Q93M28_bartonella
70	35	77.8	112	2	Q93N03_9RHIZ	Q93N03_bartonella
71	35	77.8	112	2	Q9RB93_9RHIZ	Q9RB93_bartonella
72	35	77.8	141	2	Q66330_9BACT	Q66330_unidentifie
73	35	77.8	175	2	Q4XBB9_PLACH	Q4XBB9_plasmodium
74	35	77.8	186	2	Q6ZEV6_SYNY3	Q6ZEV6_synechocyst
75	35	77.8	206	2	Q9RMP8_BARTA	Q9RMP8_bartonella
76	35	77.8	240	2	Q82S24_NITETU	Q82S24_nitrosomona
77	35	77.8	273	2	Q827W6_STRAW	Q827W6_streptomyce
78	35	77.8	291	2	Q49577_ARATH	Q49577_arabidopsi
79	35	77.8	294	2	Q8GMR8_ARATH	Q8GMR8_arabidopsi
80	35	77.8	300	2	Q5VHV5_9RHIZ	Q5VHV5_bartonella
81	35	77.8	300	2	Q5VHV8_9RHIZ	Q5VHV8_bartonella
82	35	77.8	307	2	Q6J6V4_9RHIZ	Q6J6V4_bartonella
83	35	77.8	309	2	Q8G9E7_BARTH	Q8G9E7_bartonella
84	35	77.8	316	2	Q4Y728_PLACH	Q4Y728_plasmodium
85	35	77.8	317	2	Q8PI06_XANAC	Q8PI06_xanthomonas
86	35	77.8	321	2	Q59237_9RHIZ	Q59237_bartonella
87	35	77.8	322	1	CISY_BARTA	Q59237_bartonella
88	35	77.8	330	2	Q6J6W0_BARTA	Q6J6W0_bartonella
89	35	77.8	330	2	Q6J6V1_BARTA	Q6J6V1_bartonella
90	35	77.8	331	2	Q9S306_9RHIZ	Q9S306_bartonella
91	35	77.8	338	2	Q699Y2_9RHIZ	Q699Y2_bartonella
92	35	77.8	345	2	Q6XPZ8_FUNHE	Q6XPZ8_fundulus he
93	35	77.8	374	1	FOS_TETFL	Q91496_tetradodon f
94	35	77.8	388	2	Q7U5L4_SYNPX	Q7U5L4_synechococc
95	35	77.8	414	2	Q631I6_BURPK	Q631I6_burkholderi
96	35	77.8	421	2	Q4SVT3_TETNG	Q4SVT3_tetradodon n
97	35	77.8	427	2	Q7N6M2_PHOIL	Q7N6M2_photobabdu
98	35	77.8	428	2	Q6D7G9_ERWCT	Q6D7G9_erwinia car
99	35	77.8	431	1	CISY_BARTH	P51033_bartonella
100	35	77.8	438	1	Q7UM14_RHOBA	Q7UM14_rhodospirill

ALIGNMENTS

```

RESULT 1
O6XUH7_9NOST PRELIMINARY; PRT; 503 AA.
ID O6XUH7_9NOST PRELIMINARY; PRT; 503 AA.
AC O6XUH7;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE DnaE (Fragment).
GN Name=dnaE;
OS Aphaniizomenon ovalisporum.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Aphaniizomenon.
OX NCBI_TaxId=75655;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=14651639; DOI=10.1046/j.1365-2958.2003.03825.x;
RA Caspi J., Amitai G., Belenkiy O., Petrokovski S.;
RT "Distribution of split DnaE inteins in cyanobacteria.";
RL Mol. Microbiol. 50:1569-1577(2003).
EMBL; AY209003; AAP47636.1; -; Genomic DNA.
DR GO; GO:0003887; F:DNA-directed DNA polymerase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR InterPro; IPR011708; Dna_pol3_alpha.
DR InterPro; IPR003587; Hedgehog_hint_N.
DR Pfam; PF07733; DNA_pol3_alpha; 1.
DR SMART; TIGR01445; Intein_Nterm; 1.
DR TIGRFAMs; TIGR01445; Intein_Nterm; 1.
DR PROSITE; PS50817; INTEIN_N_TER; 1.
FT NON_TER 1
SQ SEQUENCE 503 AA; 56941 MW; E6A35D127D63D066 CRC64;
Query Match 100.0%; Score 45; DB 2; Length 503;
Best Local Similarity 66.7%; Pred. No. 7.3;
Matches 8; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 CLXXTEIXTIVE 12
DB 403 CLSADTEITLIVE 414

RESULT 2
O6XUH5_APHHA PRELIMINARY; PRT; 712 AA.
ID O6XUH5_APHHA PRELIMINARY; PRT; 712 AA.
AC O6XUH5;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE DnaE (Fragment).
GN Name=dnaE;
OS Aphanothece halophytica.
OC Bacteria; Cyanobacteria; Chroococcales; Aphanothece.
OX NCBI_TaxId=72020;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=14651639; DOI=10.1046/j.1365-2958.2003.03825.x;
RA Caspi J., Amitai G., Belenkiy O., Petrokovski S.;
RT "Distribution of split DnaE inteins in cyanobacteria.";
RL Mol. Microbiol. 50:1569-1577(2003).
EMBL; AY209005; AAP47638.1; -; Genomic DNA.
DR GO; GO:0005737; C:Cytoplasm; IEA.
DR GO; GO:0008408; F:3'-5' exonuclease activity; IEA.
DR GO; GO:0003889; F:alpha DNA polymerase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR GO; GO:0015539; P:intra-mediated protein splicing; IEA.
DR InterPro; IPR011708; Dna_pol3_alpha.
DR InterPro; IPR003587; Hedgehog_hint_N.
DR InterPro; IPR004141; Intein_S.
DR InterPro; IPR004013; PHP_C.
DR InterPro; IPR004805; PolC_alpha.

```

```

DR Pfam; PF07733; Dna_pol3_alpha; 1.
DR Pfam; PF02811; PHP; 1.
DR SMART; SM00306; HintN; 1.
DR TIGRFAMs; TIGR01445; Intein_Nterm; 1.
DR TIGRFAMs; TIGR00594; PolC; 1.
DR PROSITE; PS50817; INTEIN_N_TER; 1.
FT NON_TER 1
SQ SEQUENCE 712 AA; 81509 MW; F02652B05B0AE4D1 CRC64;
Query Match 100.0%; Score 45; DB 2; Length 712;
Best Local Similarity 66.7%; Pred. No. 10;
Matches 8; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 CLXXTEIXTIVE 12
DB 603 CLSYDEITLIVE 614

RESULT 3
DPO3A_SYNY3 STANDARD; PRT; 1355 AA.
ID DPO3A_SYNY3 STANDARD; PRT; 1355 AA.
AC P74750; P73215;
DT 15-DEC-1998 (Rel. 37, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE DNA polymerase III alpha subunit (EC 2.7.7.7) [Contains: Ssp dnaE
intein].
GN Name=dnaE-N; Ordered locus names=slr0603;
GN and
OS Synechocystis sp. (strain PCC 6803);
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxId=1148;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirose M., Sugita M., Saito K., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naito K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
Synechocystis sp. strain PCC6803. II. Sequence determination of the
entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
RN [2]
RP CHARACTERIZATION OF SPLIT INTEIN.
RX MEDLINE=98356139; PubMed=9689062; DOI=10.1073/pnas.95.16.9226;
RA Wu H., Hu Z., Liu X.-Q.;
RT "Protein trans-splicing by a split intein encoded in a split DnaE gene
of Synechocystis sp. PCC6803.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:9226-9231(1998).
RN [3]
RP CHARACTERIZATION OF SPLIT INTEIN.
RX MEDLINE=20200384; PubMed=10734038; DOI=10.1074/jbc.275.13.9091;
RA Evans T.C. Jr., Martin D., Kolly R., Paine D., Sun L., Ghosh I.,
RA Chen L., Benner J., Liu X.-Q., Xu M.-Q.;
RT "Protein trans-splicing and cyclization by a naturally split intein
from the dnaE gene of Synechocystis species PCC6803.";
RL J. Biol. Chem. 275:9091-9094(2000).
RN [4]
RP CHARACTERIZATION OF SPLIT INTEIN.
RX MEDLINE=21093115; PubMed=11170467; DOI=10.1021/bi001786g;
RA Martin D.D., Xu M.-Q., Evans T.C. Jr.;
RT "Characterization of a naturally occurring trans-splicing intein from
Synechocystis sp. PCC6803.";
RL Biochemistry 40:1393-1402(2001).
CC -I- FUNCTION: DNA polymerase III is a complex, multichain enzyme
responsible for most of the replicative synthesis in bacteria.
This DNA polymerase also exhibits 3' to 5' exonuclease activity.
CC The alpha chain is the DNA polymerase (by similarity).
CC -I- CATALYTIC ACTIVITY: Deoxyribonucleoside triphosphate + DNA(n) =
diphosphate + DNA(n+1).

```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocelebration Ltd.

OM protein - protein search, using sw model

Run on: May 4, 2006, 03:16:20 ; Search time 253.333 Seconds
(without alignments)
20.813 Million cell updates/sec

Title: US-10-799-326-77

Perfect score: 45

Sequence: 1 CLXXXTREIXTVE 12

Scoring table: BIOSUM62DX

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

A_Geneseq 21: *
1: geneseqp1990s: *
2: geneseqp1990s: *
3: geneseqp2000s: *
4: geneseqp2001s: *
5: geneseqp2002s: *
6: geneseqp2003as: *
7: geneseqp2003as: *
8: geneseqp2004s: *
9: geneseqp2005s: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	100.0	64	3	AAB23745 Ssp DnaE
2	45	100.0	123	5	ABD98433 N-Intein
3	45	100.0	123	7	ABW02089 Synchocy
4	45	100.0	123	7	ABW02087 Synchocy
5	45	100.0	123	7	ADL01661 Synchocy
6	45	100.0	123	7	ADL01639 Synchocy
7	45	100.0	124	4	ABG99084 Synchocy
8	45	100.0	124	8	ADL23818 Synchocy
9	45	100.0	326	7	ADL01676 GFP prote
10	45	100.0	416	7	ADJ71769 GFP prote
11	45	100.0	432	7	ADJ71748 Aptermer P
12	42	93.3	53	7	ABP04271 Human ORF
13	40	88.9	817	4	ABB60368 Drosophil
14	39	86.7	399	4	ABB75224 Drosophil
15	39	86.7	399	5	AAE30522 Fruit fly
16	39	86.7	426	6	AAO30086 Yersinia
17	39	86.7	426	6	ABU50261 Protein e
18	39	86.7	427	6	ABU40781 Protein e
19	39	86.7	427	7	ADP04946 Bacteri
20	39	86.7	914	4	ABB68295 Drosophil
21	38	84.4	306	7	ADP07209 Bacteri
22	36	80.0	20	4	AAE02625 Plasmid P
23	36	80.0	65	3	AAE42958 Arabidops
24	36	80.0	66	3	AAE42957 Arabidops

25	36	80.0	426	8	ADN25299 Bacterial
26	36	80.0	427	8	ADN17741 Bacterial
27	36	80.0	479	8	ADS30015 Bacterial
28	36	80.0	526	6	ABU26990 Protein e
29	36	80.0	553	2	AA35682 C. pneumo
30	36	80.0	714	8	ADS21016 Bacterial
31	36	80.0	783	8	ADS41748 Bacterial
32	36	80.0	1184	6	ABU25573 Protein e
33	35	77.8	62	4	AAI17833 Peptide #
34	35	77.8	62	4	ABB36852 Peptide #
35	35	77.8	62	4	AAI17833 Peptide #
36	35	77.8	62	4	AAI17833 Peptide #
37	35	77.8	62	4	AAI17833 Peptide #
38	35	77.8	62	4	AAI17833 Peptide #
39	35	77.8	62	4	AAI17833 Peptide #
40	35	77.8	62	4	AAI17833 Peptide #
41	35	77.8	62	4	AAI17833 Peptide #
42	35	77.8	62	4	AAI17833 Peptide #
43	35	77.8	62	4	AAI17833 Peptide #
44	35	77.8	62	4	AAI17833 Peptide #
45	35	77.8	62	4	AAI17833 Peptide #
46	35	77.8	62	4	AAI17833 Peptide #
47	35	77.8	62	4	AAI17833 Peptide #
48	35	77.8	62	4	AAI17833 Peptide #
49	35	77.8	62	4	AAI17833 Peptide #
50	35	77.8	62	4	AAI17833 Peptide #
51	35	77.8	62	4	AAI17833 Peptide #
52	35	77.8	62	4	AAI17833 Peptide #
53	35	77.8	62	4	AAI17833 Peptide #
54	35	77.8	62	4	AAI17833 Peptide #
55	35	77.8	62	4	AAI17833 Peptide #
56	35	77.8	62	4	AAI17833 Peptide #
57	35	77.8	62	4	AAI17833 Peptide #
58	35	77.8	62	4	AAI17833 Peptide #
59	35	77.8	62	4	AAI17833 Peptide #
60	35	77.8	62	4	AAI17833 Peptide #
61	35	77.8	62	4	AAI17833 Peptide #
62	35	77.8	62	4	AAI17833 Peptide #
63	35	77.8	62	4	AAI17833 Peptide #
64	35	77.8	62	4	AAI17833 Peptide #
65	35	77.8	62	4	AAI17833 Peptide #
66	35	77.8	62	4	AAI17833 Peptide #
67	35	77.8	62	4	AAI17833 Peptide #
68	35	77.8	62	4	AAI17833 Peptide #
69	35	77.8	62	4	AAI17833 Peptide #
70	35	77.8	62	4	AAI17833 Peptide #
71	35	77.8	62	4	AAI17833 Peptide #
72	35	77.8	62	4	AAI17833 Peptide #
73	35	77.8	62	4	AAI17833 Peptide #
74	35	77.8	62	4	AAI17833 Peptide #
75	35	77.8	62	4	AAI17833 Peptide #
76	35	77.8	62	4	AAI17833 Peptide #
77	35	77.8	62	4	AAI17833 Peptide #
78	35	77.8	62	4	AAI17833 Peptide #
79	35	77.8	62	4	AAI17833 Peptide #
80	35	77.8	62	4	AAI17833 Peptide #
81	35	77.8	62	4	AAI17833 Peptide #
82	35	77.8	62	4	AAI17833 Peptide #
83	35	77.8	62	4	AAI17833 Peptide #
84	35	77.8	62	4	AAI17833 Peptide #
85	35	77.8	62	4	AAI17833 Peptide #
86	35	77.8	62	4	AAI17833 Peptide #
87	35	77.8	62	4	AAI17833 Peptide #
88	35	77.8	62	4	AAI17833 Peptide #
89	35	77.8	62	4	AAI17833 Peptide #
90	35	77.8	62	4	AAI17833 Peptide #
91	35	77.8	62	4	AAI17833 Peptide #
92	35	77.8	62	4	AAI17833 Peptide #
93	35	77.8	62	4	AAI17833 Peptide #
94	35	77.8	62	4	AAI17833 Peptide #
95	35	77.8	62	4	AAI17833 Peptide #
96	35	77.8	62	4	AAI17833 Peptide #
97	35	77.8	62	4	AAI17833 Peptide #

Zheng, L.
101799326 Page 1
Seq. IDs 77 4 78

98 34 75.6 435 7 ADF75265 Thale cre
99 34 75.6 435 8 ADO01785 Thalecres
100 34 75.6 464 7 ABO63104 Kiebsieil

ALIGNMENTS

RESULT 1

ID AAB23745 standard; protein; 64 AA.

AC AAB23745;

DT 06-AUG-2003 (revised)

DT 10-JAN-2001 (first entry)

DE Ssp DnaE intein amino acid sequence.

KW Intein; Int-n; Int-c; DnaE; DnaB; encryption; trait; split gene.

OS Synchocystis sp. PCC 6308.

PN WO200052146-A2.

PD 08-SEP-2000.

PF 03-MAR-2000; 2000WO-US005448.

PR 05-MAR-1999; 99US-0122943P.

PR 02-JUL-1999; 99US-0144299P.

PR 10-NOV-1999; 99US-0164617P.

PR 10-NOV-1999; 99US-0164618P.

PA (MAXY-) MAXYGEN INC.

PI Paten PA, Lassner M, Yamamoto T, Carr B, Ness JE, Bermudez ER;

XX WPI; 2000-549566/50.

XX Unencrypting traits using splice gene sequences.

XX PS Disclosure; Fig 6B; 77pp; English.

CC The present invention describes methods of unencrypting trait encrypted

CC gene sequences to provide unencrypted RNA and proteins. The present

CC sequence represents an intein amino acid sequence which is used in the

CC simplification of the present invention. (Updated on 06-AUG-2003 to

CC correct OS field.)

XX Sequence 64 AA;

QY 1 CLXXXXTEIXIVE 12

DB 1 CLSFGTEILTIVE 12

RESULT 2

ID ABB98433 standard; peptide; 123 AA.

AC ABB98433;

DT 29-AUG-2003 (revised)

DT 17-OCT-2002 (first entry)

XX N-intein encoded by DnaE gene.

XX N-intein; DnaE; antibacterial; antiparasitic; immunosuppressive;

KW fungicide; anti-HIV; bispecific molecule; C3b-like receptor;

KW protein trans-splicing; C-intein; haematopoietic; pathogenic antigen;

KW pathogenic disease.

OS Synchocystis sp. PCC 6803.

PN WO200246208-A2.

PD 13-JUN-2002.

PF 01-NOV-2001; 2001WO-US045653.

PR 01-NOV-2000; 2000US-0244811P.

PA (ELUS-) ELUSYS THERAPEUTICS INC.

PI Himawan J;

DR WPI; 2002-599455/64.

PT Producing bispecific molecule having first antigen recognition portion

PT that binds C3b-like receptor and second antigen recognition portion

PT binding pathogenic antigenic molecule by using protein transplicing

PT technique.

PS Claim 4; Fig 3; 109pp; English.

XX The present invention relates to a method for producing a bispecific

XX molecule, which has a first antigen recognition portion (ARP) that binds

XX a C3b-like receptor, and a second ARP that binds a pathogenic antigenic

XX molecule, by using a protein trans-splicing technique. The first ARP is

XX conjugated to the N-terminus of an N-intein, and the second ARP is

XX conjugated to the C-terminus of an C-intein. The N-intein first ARP and

XX trans-splicing to produce the bispecific molecule. The bispecific

XX molecule of the invention binds to haematopoietic cells expressing a C3b-

XX like receptor on their surface and clears pathogenic antigens or

XX autoantibodies from circulation, and so can be used for treating

XX pathogenic diseases. The pathogen may be an autoimmune agent, infectious

XX agent, HIV-1, bacterium e.g., Bacillus anthracis, fungus, parasite such

XX as a protozoan, or a toxin. The present sequence is N-intein (internal

XX PROTEIN domain), encoded by the DnaE gene from Synchocystis sp.

XX PCC6803., which was used to construct the bispecific molecules of the

XX invention. (Updated on 29-AUG-2003 to standardise OS field)

XX Sequence 123 AA;

QY 1 CLXXXXTEIXIVE 12

DB 1 CLSFGTEILTIVE 12

RESULT 3

ID AEM02089 standard; protein; 123 AA.

AC AEM02089;

DT 12-FEB-2004 (first entry)

DE Synchocystis sp. PCC6803 split intein Ssp DnaE Int-n, Pint-n.

KW Inactive recombinase element; promoter; ssr; site-specific recombinase;

KW Int; split intein; trans-protein splicing.

OS Synchocystis sp.

PN US2003194809-A1.

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: May 4, 2006, 03:16:20 ; Search time 316.667 Seconds
(without alignments)
20.813 Million cell updates/sec

Title: US-10-799-326-78
Perfect score: 58
Sequence: 1 DGXIRARXXHXXFT 15

Scoring table: ELOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 43978781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 100 summaries

Database :

A_Geneseq_21: *
1: geneseqp1980s: *
2: geneseqp1990s: *
3: geneseqp2000s: *
4: geneseqp2001s: *
5: geneseqp2002s: *
6: geneseqp2003as: *
7: geneseqp2003bs: *
8: geneseqp2004s: *
9: geneseqp2005s: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	58	100.0	123	5	ABP98433 N-Intein
2	58	100.0	123	7	ABW02089 Synchocy
3	58	100.0	123	7	ABW02087 Synchocy
4	58	100.0	123	7	ADL01661 Synchocy
5	58	100.0	123	7	ADL01639 Synchocy
6	58	100.0	124	4	ABG99084 Synchocy
7	58	100.0	124	8	ADL23818 Synchocy
8	58	100.0	320	7	ADL01677 Synchocy
9	58	100.0	326	7	ADL01676 Synchocy
10	58	100.0	416	7	ADJ71769 GFP prote
11	58	100.0	432	7	ADJ71748 Aptamer P
12	58	100.0	608	4	ABG99139 Pyrococu
13	46	79.3	610	4	ABG99096 Pyrococu
14	46	79.3	610	8	ADL23830 Pyrococu
15	46	79.3	1437	4	ABP96067 Putative
16	46	79.3	1437	8	ADN18584 Bacteri
17	45	77.6	154	8	ADU48592 Synchocy
18	45	77.6	155	4	ABG99117 Synchocy
19	45	77.6	162	8	ADL90459 Intein am
20	45	77.6	198	4	ABG99099 Modified
21	45	77.6	198	4	ADL23833 Synchocy
22	45	77.6	199	4	ABG99101 Synchocy
23	45	77.6	199	8	ADL23836 Synchocy
24	45	77.6	404	8	ADL90496 Retrovira

25	45	77.6	408	8	ADH08482 Retrovira
26	45	77.6	431	4	ABG99075 Synchocy
27	45	77.6	431	8	ADL23809 Synchocy
28	45	77.6	434	4	ABG99111 N-termina
29	45	77.6	434	8	ADL23856 N-termina
30	45	77.6	437	4	ABG99112 Synchocy
31	45	77.6	437	8	ADL23857 DMB inte
32	45	77.6	438	4	ABG99102 Synchocy
33	45	77.6	438	4	ABG99103 Synchocy
34	45	77.6	438	4	ABG99107 Synchocy
35	45	77.6	438	4	ABG99108 Synchocy
36	45	77.6	438	4	ABG99106 Synchocy
37	45	77.6	438	4	ABG99109 Synchocy
38	45	77.6	438	4	ABG99110 Synchocy
39	45	77.6	438	4	ABG99105 Synchocy
40	45	77.6	438	4	ABG99104 Synchocy
41	45	77.6	438	8	ADL23846 Intein mu
42	45	77.6	438	8	ADL23842 Intein mu
43	45	77.6	438	8	ADL23848 Intein mu
44	45	77.6	438	8	ADL23844 Intein mu
45	45	77.6	438	8	ADL23840 Intein mu
46	45	77.6	438	8	ADL23838 Synchocy
47	45	77.6	438	8	ADL23850 Intein mu
48	45	77.6	438	8	ADL23852 Intein mu
49	45	77.6	438	8	ADL23854 Intein mu
50	45	77.6	949	8	ADL90350 Clostridi
51	45	77.6	1420	8	ADL90392 Clostridi
52	45	77.1	1390	8	ADN46970 Thermococ
53	42	72.4	434	6	ABU34740 Protein e
54	42	72.4	434	6	ABU37031 Protein e
55	42	72.4	550	9	ABE91490 Microbia
56	42	72.4	561	3	AA57958 Rice dimi
57	42	72.4	561	7	ABM87491 Rice abio
58	42	72.4	561	8	ADL45403 Rice isop
59	42	72.4	561	8	ADK14874 Rice Osdi
60	42	72.4	561	8	ADT59956 Plant pol
61	42	72.4	1108	8	ADN18594 Bacteri
62	42	72.4	1108	8	ADS41884 Bacteri
63	41	70.7	91	8	ADO22000 Clostridi
64	41	70.7	136	8	ADO21977 Clostridi
65	41	70.7	136	8	ADO21973 Clostridi
66	41	70.7	482	8	ADY09653 Plant ful
67	41	70.7	648	6	ABU21746 Protein e
68	41	70.7	647	6	ABU26446 Protein e
69	41	70.7	1847	3	AA51631 M. jannas
70	41	70.7	1847	3	AA52002 M. jannas
71	41	70.7	1847	8	ADN18422 Bacteri
72	40	69.0	136	8	ADO21975 Clostridi
73	40	69.0	136	8	ADO21976 Clostridi
74	40	69.0	136	8	ADO21974 Clostridi
75	40	69.0	136	8	ADO21983 Clostridi
76	40	69.0	144	8	ADO21982 Clostridi
77	40	69.0	152	6	ABU00496 S. pneumo
78	40	69.0	152	6	ABP81304 Streptoco
79	40	69.0	152	8	ADK48299 Streptoco
80	40	69.0	152	8	ADM92064 S. pneumo
81	40	69.0	184	4	ABG18844 Novel hum
82	40	69.0	209	8	ADR48422 Novel S.
83	40	69.0	209	8	AEAS8712 Streptoco
84	40	69.0	811	2	AAW10570 Bacteri
85	40	69.0	2404	4	ABB69209 Drosophil
86	39	67.2	98	4	AAU61677 Propionib
87	39	67.2	98	6	ABM58196 Propionib
88	39	67.2	129	8	ADO21988 Mycobacte
89	39	67.2	147	4	ABG99085 Mycobacte
90	39	67.2	147	8	ADL23819 Mycobacte
91	39	67.2	169	4	AAW06800 Human foe
92	39	67.2	171	7	ADM25648 Hyperther
93	39	67.2	379	3	ABM15998 E. coli p
94	39	67.2	448	6	ABU45015 Protein e
95	39	67.2	579	9	AEAO6490 A. rolfsi
96	39	67.2	579	9	ADM43113 A. rolfsi
97	39	67.2	579	9	ABE72806 Glucoamy1

98 39 67.2 604 6 ABU36131 Protein e
99 39 67.2 666 4 ABB60239 Drosophila
100 39 67.2 827 6 ABU33836 Protein e

ALIGNMENTS

RESULT 1

AB98433
ID ABB98433 standard; peptide; 123 AA.

AC ABB98433;

DT 29-AUG-2003 (revised)

DT 17-OCT-2002 (first entry)

DE N-intein encoded by DnaE gene.

XX N-intein; DnaE; antibacterial; antiparasitic; immunosuppressive;

KM fungicide; anti-HIV; bispecific molecule; C3b-like receptor;

KM protein trans-splicing; C-intein; haematopoietic; pathogenic antigen;

KM pathogenic disease.

XX Synecchocystis sp. PCC 6803.

XX WO200246208-A2.

XX 13-UN-2002.

XX 01-NOV-2001; 2001WO-US045653.

XX 01-NOV-2000; 2000US-0244811P.

XX (ELUS-) ELUSYS THERAPEUTICS INC.

XX Himawan J;

XX WPI; 2002-599455/64.

PT Producing bispecific molecule having first antigen recognition portion
PT that binds C3b-like receptor and second antigen recognition portion
PT binding pathogenic antigenic molecule by using protein trans-splicing
PT technique.

PS Claim 4; Fig 3; 109pp; English.

CC The present invention relates to a method for producing a bispecific
CC molecule, which has a first antigen recognition portion (ARP) that binds
CC a C3b-like receptor, and a second ARP that binds a pathogenic antigenic
CC molecule, by using a protein trans-splicing technique. The first ARP is
CC conjugated to the N-terminus of an N-intein, and the second ARP is
CC conjugated to the C-terminus of an C-intein. The N-intein first ARP and
CC trans-splicing to produce the bispecific molecule. The bispecific
CC molecule of the invention binds to haematopoietic cells expressing a C3b-
CC like receptor on their surface and clears pathogenic antigens or
CC autoantibodies from circulation, and so can be used for treating
CC pathogenic diseases. The pathogen may be an autoimmune agent, infectious
CC agent, HIV-1, bacterium e.g., Bacillus anthracis, fungus, parasite such
CC as a protozoan, or a toxin. The present sequence is N-intein (Internal
CC protein domain), encoded by the DnaE gene from Synecchocystis sp.
CC PCC6803, which was used to construct the bispecific molecules of the
CC invention. (Updated on 29-AUG-2003 to standardise OS field)

XX Sequence 123 AA;

Query Match 100.0%; Score 58; DB 5; Length 123;

Best Local Similarity 66.7%; Pred. No. 0.051;

Matches 10; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

1 DGXVIRATXXHXEXT 15
||:|||||:|:|:|

Db 62 DGSVIRATSDHRLT 76

RESULT 2

ABW02089
ID ABW02089 standard; protein; 123 AA.

AC ABW02089;

DT 12-FEB-2004 (first entry)

DE Synecchocystis sp. PCC6803 split intein Ssp DnaE int-n, Plnt-n.

XX Inactive recombinase element; promoter; ssr; site-specific recombinase;

KM int; split intein; trans-protein splicing.

XX Synecchocystis sp.

XX US2003194809-A1.

XX 16-OCT-2003.

XX 29-JAN-2003; 2003US-00353454.

XX 11-APR-2002; 2002US-0352798P.

XX (YADA/) YADAV N S.

XX (YANG/) YANG J.

XX Yadav NS, Yang J;

XX WPI; 2003-844461/78.

XX N-PSDB; AAD63697.

PT New inactive recombinase element comprising DNA constructs having general
PT structure in 5' to 3' orientation, useful for conditionally activating or
PT removing a trait gene in a hybrid plant.

XX Example 1; Page 30; 0pp; English.

CC The invention relates to an inactive recombinase element comprising a DNA
CC construct having the general structure in 5' to 3' orientation comprising
CC a DNA construct having a general structure in 5' to 3' end orientation
CC comprising P1 (promoter)-ssrN (N-terminal portion of a site-specific
CC recombinase)-intN (N-terminal portion of a split intein) or a DNA
CC construct having a general structure in 5' to 3' end orientation
CC comprising P2 (promoter)-intC (C-terminal portion of a split intein)-ssrC
CC (the C-terminal portion of a site-specific recombinase) where the
CC inactive recombinase elements of both the DNA constructs when present
CC together in a cell will form an inactive recombinase protein by intein-
CC mediated trans-protein splicing. The inactive recombinase element, system
CC and methods are useful for conditionally activating or removing a trait
CC gene in a hybrid plant. The present sequence is Synecchocystis sp. PCC6803
CC split intein Ssp DnaE int-n

XX Sequence 123 AA;

Query Match 100.0%; Score 58; DB 7; Length 123;

Best Local Similarity 66.7%; Pred. No. 0.051;

Matches 10; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

1 DGXVIRATXXHXEXT 15
||:|||||:|:|:|

62 DGSVIRATSDHRLT 76

Query Match 100.0%; Score 58; DB 5; Length 123;

Best Local Similarity 66.7%; Pred. No. 0.051;

Matches 10; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

12-FEB-2004 (first entry)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: May 4, 2006, 03:32:22 ; Search time 29.3333 Seconds
(without alignments)
33.822 Million cell updates/sec

Title: US-10-799-326-77
Perfect score: 45
Sequence: 1 CLXXXIXIVE 12

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/1aa/5-COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/6-COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/H-COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/PCTUS-COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/RE-COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	100.0	124	2	US-09-800-170-10
2	39	86.7	427	2	US-09-543-681A-5231
3	38	84.4	306	2	US-09-543-681A-7494
4	36	80.0	20	2	US-09-687-875A-21
5	36	80.0	538	2	US-09-438-185A-1026
6	36	80.0	553	2	US-09-198-452A-1100
7	35	77.8	331	2	US-10-104-047-2976
8	35	77.8	815	2	US-10-104-047-3390
9	34	75.6	247	2	US-09-230-196-16
10	34	75.6	272	2	US-09-540-236-2133
11	34	75.6	326	2	US-09-230-196-20
12	34	75.6	334	2	US-09-438-185A-11
13	34	75.6	361	2	US-09-404-296B-30
14	34	75.6	435	2	US-09-533-028-56
15	34	75.6	456	2	US-09-248-796A-23791
16	34	75.6	464	2	US-09-489-039A-9621
17	34	75.6	810	2	US-09-198-452A-20
18	34	75.6	810	2	US-09-198-452A-1128
19	34	75.6	819	2	US-09-438-185A-1055
20	34	75.6	1205	2	US-09-330-330-1
21	33	73.3	305	1	US-08-118-270-39
22	33	73.3	305	4	PCT-US93-08528-59
23	33	73.3	313	2	US-09-438-185A-1056
24	33	73.3	320	2	US-09-198-452A-1130
25	33	73.3	343	2	US-09-109-204-32
26	33	73.3	343	2	US-09-490-032-32
27	33	73.3	343	2	US-09-464-377-7

28	73.3	369	2	US-09-949-016-7375	Sequence 7375, Ap
29	73.3	690	2	US-09-538-092-1176	Sequence 1176, Ap
30	73.3	2627	1	US-08-751-189-3	Sequence 3, Appl
31	73.3	2627	1	US-09-060-836-3	Sequence 3, Appl
32	73.3	2627	1	US-09-184-445-3	Sequence 3, Appl
33	73.3	3433	2	US-09-091-501B-10	Sequence 10, Appl
34	73.3	3433	2	US-09-538-092-1136	Sequence 1136, Ap
35	73.3	83	2	US-09-248-796A-23714	Sequence 22174, A
36	71.1	110	2	US-09-205-228-1080	Sequence 1080, Ap
37	71.1	110	2	US-10-004-860-1080	Sequence 24389, A
38	71.1	141	2	US-09-248-796A-24389	Sequence 10389, A
39	71.1	197	4	PCT-US95-03384-2	Sequence 2, Appl
40	71.1	197	4	PCT-US95-03384-2	Sequence 2, Appl
41	71.1	309	2	US-09-346-433E-37	Sequence 37, Appl
42	71.1	309	2	US-09-862-027-37	Sequence 37, Appl
43	71.1	326	2	US-09-248-796A-14157	Sequence 14157, A
44	71.1	399	2	US-09-949-016-8459	Sequence 8459, Ap
45	71.1	399	2	US-09-949-016-8460	Sequence 8460, Ap
46	71.1	399	2	US-09-949-016-8461	Sequence 8461, Ap
47	71.1	399	2	US-09-949-016-8462	Sequence 8462, Ap
48	71.1	424	2	US-09-205-228-1079	Sequence 1079, Ap
49	71.1	424	2	US-10-004-860-1079	Sequence 1079, Ap
50	71.1	436	2	US-09-134-000C-3950	Sequence 3950, Ap
51	71.1	452	1	US-08-205-719-3	Sequence 3, Appl
52	71.1	454	1	US-07-915-720D-17	Sequence 17, Appl
53	71.1	454	2	US-09-025-543-17	Sequence 17, Appl
54	71.1	461	2	US-10-142-835-35	Sequence 35, Appl
55	71.1	467	1	US-07-847-562-2	Sequence 2, Appl
56	71.1	477	1	US-08-240-328-2	Sequence 2, Appl
57	71.1	477	1	US-08-990-849-2	Sequence 2, Appl
58	71.1	477	6	5245013-3	Patent No. 5245013
59	71.1	481	1	US-08-215-089-9	Sequence 9, Appl
60	71.1	481	4	PCT-US95-03384-9	Sequence 9, Appl
61	71.1	482	2	US-09-538-092-486	Sequence 486, Ap
62	71.1	491	2	US-09-543-681A-1588	Sequence 4588, Ap
63	71.1	538	2	US-09-270-767-57575	Sequence 57575, A
64	71.1	698	2	US-09-270-767-42292	Sequence 42292, A
65	71.1	741	2	US-09-854-856-60	Sequence 60, Appl
66	71.1	741	2	US-10-010-720-60	Sequence 60, Appl
67	71.1	769	2	US-09-854-856-44	Sequence 44, Appl
68	71.1	769	2	US-10-010-720-44	Sequence 44, Appl
69	71.1	801	2	US-09-854-856-28	Sequence 28, Appl
70	71.1	801	2	US-10-010-720-28	Sequence 28, Appl
71	71.1	823	2	US-09-949-016-8339	Sequence 8339, Ap
72	71.1	829	2	US-09-854-856-12	Sequence 12, Appl
73	71.1	829	2	US-10-010-720-12	Sequence 12, Appl
74	71.1	858	2	US-07-956-483-17	Sequence 17, Appl
75	71.1	858	2	US-09-206-551-18	Sequence 18, Appl
76	71.1	867	2	US-09-248-796A-19207	Sequence 19207, A
77	71.1	894	2	US-09-854-856-54	Sequence 54, Appl
78	71.1	894	2	US-10-010-720-54	Sequence 54, Appl
79	71.1	922	2	US-09-854-856-38	Sequence 38, Appl
80	71.1	922	2	US-10-010-720-38	Sequence 38, Appl
81	71.1	940	2	US-09-540-236-2645	Sequence 2645, Ap
82	71.1	954	2	US-09-854-856-22	Sequence 22, Appl
83	71.1	954	2	US-10-010-720-22	Sequence 22, Appl
84	71.1	982	2	US-09-854-856-6	Sequence 6, Appl
85	71.1	982	2	US-10-010-720-6	Sequence 6, Appl
86	71.1	1911	2	US-09-854-856-64	Sequence 64, Appl
87	71.1	1911	2	US-10-010-720-64	Sequence 64, Appl
88	71.1	1939	2	US-09-854-856-48	Sequence 48, Appl
89	71.1	1939	2	US-10-010-720-48	Sequence 48, Appl
90	71.1	1971	2	US-09-854-856-32	Sequence 32, Appl
91	71.1	1971	2	US-10-010-720-32	Sequence 32, Appl
92	71.1	1999	2	US-09-854-856-16	Sequence 16, Appl
93	71.1	1999	2	US-10-010-720-16	Sequence 16, Appl
94	71.1	2004	2	US-09-854-856-58	Sequence 58, Appl
95	71.1	2004	2	US-10-010-720-58	Sequence 58, Appl
96	71.1	2032	2	US-09-854-856-42	Sequence 42, Appl
97	71.1	2032	2	US-10-010-720-42	Sequence 42, Appl
98	71.1	2048	2	US-09-854-856-62	Sequence 62, Appl
99	71.1	2048	2	US-10-010-720-62	Sequence 62, Appl
100	71.1	2064	2	US-09-854-856-26	Sequence 26, Appl

ALIGNMENTS

```
RESULT 1
US-09-800-170-10
; Sequence 10, Application US/09800170
; Patent No. 6481667
; GENERAL INFORMATION:
; APPLICANT: Kinseila, Todd
; TITLE OF INVENTION: IN VIVO PRODUCTION OF CYCLIC PEPTIDES
; FILE REFERENCE: A-68614-1/DJB/RMS/RMK
; CURRENT APPLICATION NUMBER: US/09/800,170
; CURRENT FILING DATE: 2001-08-28
; PRIOR APPLICATION NUMBER: US 60/187,130
; PRIOR FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Synechocystis PCC6803
US-09-800-170-10

Query Match      100.0%; Score 45; DB 2; Length 124;
Best Local Similarity 66.7%; Pred. No. 1.7;
Matches 8; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY      1 CLXXXTEIXTVE 12
        |||:|||||
Db      2 CLSFGTEILTVE 13

RESULT 2
US-09-543-681A-5231
; Sequence 5231, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NOCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5231
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-5231

Query Match      86.7%; Score 39; DB 2; Length 427;
Best Local Similarity 58.3%; Pred. No. 86;
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY      1 CLXXXTEIXTVE 12
        |||:|||||
Db      271 CLRMLEIKTVE 282

RESULT 3
US-09-543-681A-7494
; Sequence 7494, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NOCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
```

```
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7494
; LENGTH: 306
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-7494

Query Match      84.4%; Score 38; DB 2; Length 306;
Best Local Similarity 41.7%; Pred. No. 89;
Matches 5; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

QY      1 CLXXXTEIXTVE 12
        |||:|||||
Db      275 CYYNTQIPIE 286

RESULT 4
US-09-687-875A-21
; Sequence 21, Application US/09687875A
; Patent No. 6544786
; GENERAL INFORMATION:
; APPLICANT: Xiao, Xiao
; APPLICANT: Liu, Paul
; TITLE OF INVENTION: METHOD AND VECTOR FOR PRODUCING AND TRANSFERRING TRANS-SPICED PE
; FILE REFERENCE: 00792
; CURRENT APPLICATION NUMBER: US/09/687,875A
; CURRENT FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/158,868
; PRIOR FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: pXX-N4 dystrophin-intein peptide junction
US-09-687-875A-21

Query Match      80.0%; Score 36; DB 2; Length 20;
Best Local Similarity 60.0%; Pred. No. 9;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY      1 CLXXXTEIXT 10
        |||:|||||
Db      11 CLSFGTEILT 20

RESULT 5
US-09-438-185A-1026
; Sequence 1026, Application US/09438185A
; Patent No. 6822071
; GENERAL INFORMATION:
; APPLICANT: Stephens, Richard
; APPLICANT: Mitchell, Wayne
; APPLICANT: Kaiman, Sue
; APPLICANT: Davis, Ronald
; TITLE OF INVENTION: Chlamydia Pneumoniae Genome Sequence
; FILE REFERENCE: 018941-000411US
; CURRENT APPLICATION NUMBER: US/09/438,185A
; CURRENT FILING DATE: 2002-03-13
; PRIOR APPLICATION NUMBER: US 60/108,279
; PRIOR FILING DATE: 1998-11-12
; PRIOR APPLICATION NUMBER: US 60/128,606
; PRIOR FILING DATE: 1999-04-08
; NUMBER OF SEQ ID NOS: 1074
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1026
; LENGTH: 538
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocelebration Ltd.

OM protein - protein search, using sw model

Run on: May 4, 2006, 03:27:26 ; Search time 20.8889 Seconds
(without alignments)
55.273 Million cell updates/sec

Title: US-10-799-326-77
Perfect score: 45
Sequence: 1 CLXXXTEIXTIVE 12

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 283416 segs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	100.0	.897	2	S76958
2	44	97.8	877	2	AC2253
3	39	86.7	426	2	AD0136
4	37	82.2	544	2	T07593
5	36	80.0	369	2	E90532
6	36	80.0	526	2	F86618
7	36	80.0	526	2	C72005
8	36	80.0	526	2	B81533
9	36	80.0	815	2	AG2278
10	35	77.8	104	2	SS3942
11	35	77.8	291	2	T04494
12	35	77.8	431	2	I40044
13	35	77.8	443	1	G69017
14	35	77.8	443	2	T45147
15	35	77.8	451	2	A64310
16	35	77.8	571	2	A10506
17	35	77.8	748	2	AD0866
18	35	77.8	748	2	F91089
19	35	77.8	748	2	H85934
20	35	77.8	748	2	F65065
21	35	77.8	962	2	T05845
22	35	77.8	2078	2	T25400
23	34	75.6	203	2	G97380
24	34	75.6	203	2	AF2598
25	34	75.6	225	2	A40326
26	34	75.6	260	2	C69130
27	34	75.6	297	2	A72130
28	34	75.6	324	2	F69287
29	34	75.6	326	2	T19070

30	34	75.6	350	2	G97098	probable cation ef
31	34	75.6	357	2	H65041	hypothetical prote
32	34	75.6	357	2	T08642	hypothetical prote
33	34	75.6	361	2	C84613	probable cyclin D
34	34	75.6	370	2	T32615	hypothetical prote
35	34	75.6	383	2	S51651	cyclin delta-2 - A
36	34	75.6	465	2	D63324	UDP-glucose dehydr
37	34	75.6	475	2	AC1246	branched-chain alp
38	34	75.6	475	2	AG1608	hypothetical prote
39	34	75.6	486	2	T70154	replication initia
40	34	75.6	522	2	A81537	hypothetical prote
41	34	75.6	748	2	AG0096	phosphoenolpyruvat
42	34	75.6	774	2	B86492	hypothetical prote
43	34	75.6	774	2	H81540	hypothetical prote
44	34	75.6	811	2	E72003	hypothetical prote
45	34	75.6	881	2	T31818	hypothetical prote
46	34	75.6	1095	2	T41171	importin beta subu
47	33	73.3	82	2	I50984	MHC class II beta
48	33	73.3	82	2	I50986	MHC class II beta
49	33	73.3	82	2	I50968	MHC class II beta
50	33	73.3	82	2	I51147	MHC class II beta
51	33	73.3	138	2	T25698	hypothetical prote
52	33	73.3	265	2	T27673	hypothetical prote
53	33	73.3	276	2	D72001	hypothetical prote
54	33	73.3	276	2	D86622	hypothetical prote
55	33	73.3	348	1	OOHUB	opsin, blue-sensit
56	33	73.3	352	2	A81861	impr3/mucB/samb fam
57	33	73.3	429	2	S41527	citrate (6i)-synth
58	33	73.3	436	2	JN0591	serotonin receptor
59	33	73.3	445	2	I46225	mucin A, tracheal
60	33	73.3	597	2	B53978	protein-tyrosine-p
61	33	73.3	678	2	H96552	hypothetical prote
62	33	73.3	694	2	A53978	protein-tyrosine-p
63	33	73.3	831	2	T25827	hypothetical prote
64	33	73.3	1216	2	H84629	hypothetical prote
65	33	73.3	1319	2	H84542	hypothetical prote
66	33	73.3	1349	2	S51471	killer toxin insen
67	33	73.3	2895	2	H85362	hypothetical prote
68	33	73.3	3433	1	S28381	uticofin - human
69	33	73.3	82	2	I51137	MHC class II beta
70	32	71.1	104	2	T30628	hypothetical prote
71	32	71.1	104	2	S41541	hypothetical prote
72	32	71.1	212	2	A81175	GTP-pyrophosphokin
73	32	71.1	217	2	AD1624	O-methyltransferas
74	32	71.1	221	2	AF1342	hypothetical prote
75	32	71.1	225	2	A31583	C-reactive protein
76	32	71.1	243	2	S73698	hypothetical prote
77	32	71.1	255	2	A64723	hypothetical prote
78	32	71.1	262	2	T32979	hypothetical prote
79	32	71.1	324	2	AH3134	hypothetical prote
80	32	71.1	350	2	C36769	13R protein - huma
81	32	71.1	353	2	T44027	hypothetical prote
82	32	71.1	355	2	D98153	oligopeptidase ABC t
83	32	71.1	431	2	B82119	citrate synthase V
84	32	71.1	456	2	AH0240	probable chlorophd
85	32	71.1	456	2	T46986	hypothetical prote
86	32	71.1	457	2	A60966	hypothetical integ
87	32	71.1	457	2	A64924	probable membrane
88	32	71.1	457	2	H85773	transcription init
89	32	71.1	457	2	D90925	probable citrate s
90	32	71.1	459	2	A83878	acetoaldehyde dehydr
91	32	71.1	461	2	T09341	hydroxymethylgluta
92	32	71.1	475	2	I76668	pregnane-specific
93	32	71.1	477	2	A35843	lipopolysaccharide
94	32	71.1	482	2	S37845	transcription init
95	32	71.1	509	2	C84858	probable citrate s
96	32	71.1	570	2	T08778	hypothetical prote
97	32	71.1	574	2	T29615	hypothetical prote
98	32	71.1	677	2	H86208	protein F22G5.26 (
99	32	71.1	689	2	T08988	cadmium-transporti
100	32	71.1	705	2	A71211	probable phosphori

ALIGNMENTS

RESULT 1

S76958

C/Directed DNA polymerase (EC 2.7.7.7) III alpha chain split extein-N containing prech
C/Species: Synecchocystis sp.

C/Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 12-Jul-2004

C/Accession: S76958

R/Kaneho, T.; Saco, S.; Korani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;

O. K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996

A/Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecchocystis

A/Reference number: S74322; MUID:37061201; PMID:8905231

A/Accession: S76958

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-897 <KAN>

A/Cross-references: UNIPARC:UPI0000164C87; EMBL:D90917; GB:AB001339; NID:91653836; PIDN:

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

R/Hu, H.; Hu, Z.; Lu, X.O.

Proc. Natl. Acad. Sci. U.S.A. 95, 9226-9231, 1998

A/Title: Protein trans-splicing by a split intem encoded in a split DnaE gene of Synecch

A/Reference number: A59016; MUID:98356139; PMID:9689062

A/Contents: annotation

C/Comment: For the other split extein precursor see PIR:S75328. For the mature spliced f

C/Genetics:

A/Gene: dnaE-n

C/Keywords: DNA biosynthesis, DNA replication, nucleotidyltransferase, protein splicing

F/1-774/Domain: DNA-directed DNA polymerase III alpha chain split extein-N #status predi

F/775-897/Domain: split intein-N #status predicted <NTN>

Query Match 100.0%; Score 45; DB 2; Length 897;

Best Local Similarity 66.7%; Pred. No. 2.2;

Matches 8; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 CLXXXTEIXIVE 12

Db 775 CLSFGTEILIVE 786

RESULT 2

AC2253

DNA polymerase III alpha chain [imported] - Nostoc sp. (strain PCC 7120)

C/Species: Nostoc sp. PCC 7120

A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C/Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 12-Jul-2004

C/Accession: AC2253

R/Kaneho, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi

Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S

DNA Res. 8, 205-213, 2001

A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A/Reference number: AB1807; MUID:21595285; PMID:11759640

A/Accession: AC2253

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-877 <KUR>

A/Cross-references: UNIPROT:Q8YR72; UNIPARC:UPI00000CE838; GB:BA000019; PIDN:BA875277.1;

A/Experimental source: strain PCC 7120

C/Genetics:

A/Gene: all3578

Query Match 97.8%; Score 44; DB 2; Length 877;

Best Local Similarity 58.3%; Pred. No. 3.4;

Matches 7; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 1 CLXXXTEIXIVE 12

Db 776 CLSYDEVLIVE 787

RESULT 3

AD0136

citrate (6i)-synthase (EC 4.1.3.7) glta [imported] - Yersinia pestis (strain CO92)

C/Species: Yersinia pestis

C/Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 05-Oct-2004

C/Accession: AD0136

R/Parkhill, J.; Wren, B.W.; Thomeon, N.R.; Titchall, R.W.; Holden, M.T.G.; Prentice, M.B.

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;

11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,

Nature 413, 523-527, 2001

A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A/Reference number: AB0001; MUID:21470413; PMID:11586360

A/Accession: AD0136

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-426 <KUR>

A/Cross-references: UNIPROT:Q8ZH07; UNIPARC:UPI00000DCAB3; GB:AL590842; PIDN:CAC89951.1;

C/Genetics:

A/Gene: gltA

C/Keywords: citrate-synthase; coenzyme A; oxo-acid-lyase

C/Superfamily: carbon-carbon lyase;

Query Match 86.7%; Score 39; DB 2; Length 426;

Best Local Similarity 58.3%; Pred. No. 17;

Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 CLXXXTEIXIVE 12

Db 271 CLKMLEIXIVE 282

RESULT 4

T07593

pectinesterase (EC 3.1.1.11) 3 precursor - tomato

N/Alternate names: pectin methylsterase

C/Species: Lycopersicon esculentum (tomato)

C/Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 09-Jul-2004

C/Accession: T07593

R/Turner, L.A.; Hartman, R.W.; Handa, A.K.

submitted to the EMBL Data Library, September 1996

A/Reference number: Z16036

A/Accession: T07593

A/Status: preliminary; translated from GB/EMBL/DD8J

A/Molecule type: DNA

A/Residues: 1-544 <TUR>

A/Cross-references: UNIPROT:Q96576; UNIPARC:UPI0000131CB; EMBL:U70676; NID:91617585; PI

A/Experimental source: strain VFWT Cherry

C/Genetics:

A/Gene: PMB3

A/Intons: 313/1; 377/1

C/Function:

A/Description: catalyzes the hydrolysis of pectin into pectate and methanol

C/Superfamily: pectinesterase

C/Keywords: carboxylic ester hydrolase

Query Match 82.2%; Score 37; DB 2; Length 544;

Best Local Similarity 50.0%; Pred. No. 55;

Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 1 CLXXXTEIXIVE 12

Db 63 CLSYSEIVITE 74

RESULT 5

E90532

gtp-binding protein [imported] - Mycoplasma pulmonis (strain UAB CTIP)

C/Species: Mycoplasma pulmonis

C/Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004

C/Accession: E90532

R/Chambaud, I.; Hellig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;

Nucleic Acids Res. 29, 2145-2153, 2001

A/Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulr

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: May 4, 2006, 03:34:13 ; Search time 15.5556 Seconds
(without alignments)
35.705 Million cell updates/sec

Title: US-10-799-326-77
Perfect score: 45
Sequence: 1 CLXXXTXITVE 12

Scoring table: BLOSUM62X
Gapop 10.0 , Gapext 0.5

Searched: 235405 seqs, 46284737 residues
Total number of hits satisfying chosen parameters: 235405

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Published Applications AA New:*
1: /SIDS/prodata/1/pubppa/US08_NEW_PUB.pep1.*
2: /SIDS/prodata/1/pubppa/US06_NEW_PUB.pep.*
3: /SIDS/prodata/1/pubppa/US07_NEW_PUB.pep.*
4: /SIDS/prodata/1/pubppa/US08_NEW_PUB.pep.*
5: /SIDS/prodata/1/pubppa/PCT_NEW_PUB.pep.*
6: /SIDS/prodata/1/pubppa/US09_NEW_PUB.pep1.*
7: /SIDS/prodata/1/pubppa/US09_NEW_PUB.pep1.*
8: /SIDS/prodata/1/pubppa/US10_NEW_PUB.pep1.*
9: /SIDS/prodata/1/pubppa/US10_NEW_PUB.pep1.*
10: /SIDS/prodata/1/pubppa/US11_NEW_PUB.pep1.*
11: /SIDS/prodata/1/pubppa/US11_NEW_PUB.pep1.*
12: /SIDS/prodata/1/pubppa/US60_NEW_PUB.pep1.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	100.0	124	11	US-11-069-642-10
2	36	80.0	369	11	US-11-087-099-2275
3	35	77.8	831	11	US-11-072-512-2976
4	35	77.8	815	11	US-11-072-512-3390
5	34	75.6	102	9	US-10-689-742-44
6	34	75.6	226	11	US-11-151-601-29
7	34	75.6	226	11	US-11-151-601-37
8	34	75.6	475	11	US-11-045-004-1004
9	33	73.3	33	9	US-10-895-064-1551
10	33	73.3	33	11	US-11-129-741-1551
11	33	73.3	165	11	US-11-096-568A-6346
12	33	73.3	325	9	US-10-506-454-1409
13	33	73.3	325	11	US-11-188-298-9894
14	33	73.3	343	8	US-10-370-959-146
15	33	73.3	353	8	US-10-370-959-150
16	33	73.3	371	8	US-10-370-959-147
17	33	73.3	385	11	US-11-096-568A-13633
18	33	73.3	423	9	US-10-506-454-937
19	32	71.1	115	11	US-11-096-568A-31408
20	32	71.1	136	11	US-11-188-298-16308
21	32	71.1	212	11	US-11-045-004-1914

22	32	71.1	221	11	US-11-045-004-2747	Sequence 2747, Ap
23	32	71.1	249	11	US-11-098-686-10924	Sequence 10924, A
24	32	71.1	382	11	US-11-096-568A-17025	Sequence 17025, A
25	32	71.1	394	11	US-11-096-568A-17024	Sequence 17024, A
26	32	71.1	423	11	US-11-096-568A-17023	Sequence 17023, A
27	32	71.1	431	11	US-11-079-463-6094	Sequence 6094, Ap
28	32	71.1	462	11	US-11-264-096-1692	Sequence 1692, Ap
29	32	71.1	704	11	US-11-096-568A-28142	Sequence 28142, A
30	32	71.1	710	11	US-11-096-568A-28141	Sequence 28141, A
31	32	71.1	774	11	US-11-096-568A-33913	Sequence 33913, A
32	32	71.1	787	11	US-11-096-568A-33912	Sequence 33912, A
33	32	71.1	802	9	US-10-330-773-590	Sequence 590, App
34	32	71.1	821	8	US-10-505-928-316	Sequence 316, App
35	32	71.1	821	11	US-11-087-227-90	Sequence 90, App1
36	32	71.1	841	11	US-11-233-510-16	Sequence 16, App1
37	32	71.1	841	11	US-11-096-568A-33911	Sequence 33911, A
38	32	71.1	873	11	US-11-096-568A-28140	Sequence 28140, A
39	32	71.1	988	9	US-10-506-454-1680	Sequence 1680, Ap
40	32	71.1	1796	11	US-11-087-099-10309	Sequence 10309, A
41	32	71.1	1796	11	US-11-188-298-9498	Sequence 9498, Ap
42	32	71.1	2204	9	US-10-495-083-8	Sequence 8, App1
43	32	71.1	2228	9	US-10-511-096-2	Sequence 2, App1
44	32	71.1	2230	9	US-10-511-096-4	Sequence 4, App1
45	32	71.1	2250	9	US-10-511-096-6	Sequence 6, App1
46	32	71.1	2252	9	US-10-511-096-8	Sequence 8, App1
47	31	68.9	28	11	US-11-235-672-42	Sequence 42, App1
48	31	68.9	40	9	US-10-957-351-15	Sequence 15, App1
49	31	68.9	74	11	US-11-096-568A-13883	Sequence 13883, A
50	31	68.9	120	9	US-10-115-609-24	Sequence 24, App1
51	31	68.9	288	8	US-10-511-937-2448	Sequence 2448, Ap
52	31	68.9	288	9	US-10-511-609-12	Sequence 12, App1
53	31	68.9	291	9	US-10-457-657-3890	Sequence 3890, Ap
54	31	68.9	306	11	US-11-096-568A-5677	Sequence 5677, Ap
55	31	68.9	351	11	US-11-096-568A-5676	Sequence 5676, Ap
56	31	68.9	367	11	US-11-087-099-3589	Sequence 3589, Ap
57	31	68.9	378	9	US-10-506-454-224	Sequence 224, App
58	31	68.9	394	11	US-11-096-568A-13356	Sequence 13356, A
59	31	68.9	405	9	US-10-763-712A-120	Sequence 120, App
60	31	68.9	434	11	US-11-045-004-892	Sequence 892, App
61	31	68.9	436	11	US-11-096-568A-5675	Sequence 5675, Ap
62	31	68.9	473	11	US-11-096-568A-13355	Sequence 13355, A
63	31	68.9	473	11	US-11-096-568A-21002	Sequence 21002, A
64	31	68.9	474	11	US-11-096-568A-21001	Sequence 21001, A
65	31	68.9	479	11	US-11-096-568A-13354	Sequence 13354, A
66	31	68.9	489	11	US-11-210-251-1	Sequence 1, App1
67	31	68.9	490	11	US-11-188-298-15737	Sequence 15737, A
68	31	68.9	490	11	US-11-188-298-18737	Sequence 18737, A
69	31	68.9	503	9	US-10-821-234-1527	Sequence 1527, Ap
70	31	68.9	524	9	US-10-821-234-1631	Sequence 1631, Ap
71	31	68.9	557	11	US-11-096-568A-21000	Sequence 21000, A
72	31	68.9	580	11	US-11-045-004-938	Sequence 938, App
73	31	68.9	671	9	US-10-194-487-346	Sequence 346, App
74	31	68.9	671	9	US-10-195-888-346	Sequence 346, App
75	31	68.9	671	9	US-10-195-888-346	Sequence 346, App
76	31	68.9	671	9	US-10-195-888-346	Sequence 346, App
77	31	68.9	676	11	US-11-135-855-28	Sequence 28, App1
78	31	68.9	717	11	US-11-135-855-29	Sequence 29, App1
79	31	68.9	843	11	US-11-072-512-2043	Sequence 2043, Ap
80	31	68.9	843	11	US-11-188-298-1848	Sequence 3848, Ap
81	31	68.9	859	11	US-11-000-463-423	Sequence 423, App
82	31	68.9	859	11	US-11-000-463-895	Sequence 895, App
83	31	68.9	1210	11	US-11-096-568A-31726	Sequence 31726, A
84	31	68.9	1254	11	US-11-096-568A-31725	Sequence 31725, A
85	31	68.9	1347	11	US-11-096-568A-31724	Sequence 31724, A
86	31	68.9	8693	11	US-11-205-109-15	Sequence 15, App1
87	31	68.9	8693	11	US-11-205-109-14	Sequence 14, App1
88	30	66.7	12	11	US-11-145-861-14	Sequence 14, App1
89	30	66.7	100	9	US-10-194-487-526	Sequence 526, App
90	30	66.7	100	9	US-10-195-888-526	Sequence 526, App
91	30	66.7	100	9	US-10-195-888-526	Sequence 526, App
92	30	66.7	120	11	US-11-112-240-2	Sequence 2, App1
93	30	66.7	120	11	US-11-112-240-2	Sequence 2, App1
94	30	66.7	120	11	US-11-112-304A-2	Sequence 2, App1

95	30	66.7	125	11	US-11-2854-096-1397	Sequence 1397, Ap
96	30	66.7	125	11	US-11-2854-096-1399	Sequence 1399, Ap
97	30	66.7	152	9	US-10-506-454-553	Sequence 553, App
98	30	66.7	213	9	US-10-242-586-64	Sequence 64, Appl
99	30	66.7	213	9	US-10-242-902-64	Sequence 64, Appl
100	30	66.7	213	9	US-10-243-116-64	Sequence 64, Appl

ALIGNMENTS

```

RESULT 1
US-11-069-642-10
; Sequence 10, Application US/11069642
; Publication No. US20050260626A1
; GENERAL INFORMATION:
; APPLICANT: LORENS, JAMES B.
; APPLICANT: PRAY, TODD R.
; APPLICANT: KINSELLA, TODD M.
; APPLICANT: BENNETT, MARK K.
; TITLE OF INVENTION: IN VIVO PRODUCTION OF CYCLIC PEPTIDES FOR
; FILE OF INVENTION: INHIBITING PROTEIN-PROTEIN INTERACTION
; TITLE REFERENCE: RIGI-022C1P3
; CURRENT APPLICATION NUMBER: US/11/069,642
; CURRENT FILING DATE: 2005-02-28
; PRIOR APPLICATION NUMBER: 10/232,758
; PRIOR FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: 10/422,536
; PRIOR FILING DATE: 2003-04-23
; PRIOR APPLICATION NUMBER: 09/800,770
; PRIOR FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: 60/187,130
; PRIOR FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 168
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Synechocystis PCC6803
; US-11-069-642-10

```

```

Query Match          100.0%;  Score 45;  DB 11;  Length 124;
Best Local Similarity 66.7%;  Pred. No. 0.21;
Matches      8;  Conservative      4;  Mismatches      0;  Gaps      0;
QY      1  CLXXXTETIXTE 12
        |||::|||::||
        db      2  CLSFGTEITVE 13

```

```

RESULT 2
US-11-087-099-2275
; Sequence 2275, Application US/11087099
; Publication No. US20060041961A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: Genes and Uses for PI
; FILE REFERENCE: 38-21(53450) B EP
; CURRENT APPLICATION NUMBER: US/11/087,099
; CURRENT FILING DATE: 2005-03-22
; NUMBER OF SEQ ID NOS: 12464
; SEQ ID NO 2275
; LENGTH: 369
; TYPE: prt
; ORGANISM: Mycoplasma pulmonis
US-11-087-099-2275

```

Query Match	80.0%;	Score 36;	DB 11;	Length 369;
Best Local Similarity	33.3%;	Pred. NO. 45;		
Matches	4;	Conservative	8;	Mismatches 0;
				Indels 0;
				Gaps 0;
QY	1	CLXXXTXIXIVE	12	
		: : : : : : :		

Db 242 CIKIETELSSIE 253

3

```

Sequence 2976, Application US/11072512
Publication No. US20060029945A1
GENERAL INFORMATION:
APPLICANT: ISOGAI, TAKAO
APPLICANT: SUGIYAMA, TOMOYASU
APPLICANT: OTSUKI, TETSUJI
APPLICANT: WAKAMATSU, AI
APPLICANT: SATO, HIROYUKI
APPLICANT: ISHII, SHIZUKO
APPLICANT: YAMAMOTO, JUN-ICHI
APPLICANT: ISONO, YUUKO
APPLICANT: HIO, YURI
APPLICANT: OTSUKA, KAORU
APPLICANT: NAGAI, KEIICHI
APPLICANT: IRIE, RYOTARO
APPLICANT: TAMECHIKA, ICHIRO
APPLICANT: SEKI, NAOHICO
APPLICANT: YOSHIKAWA, TSUTOMU
APPLICANT: OTSUKA, MOTOKYO
APPLICANT: NAGAHARI, KENJI
APPLICANT: MASUHO, YASUHIKO
TITLE OF INVENTION: Novel full length cDNA
FILE REFERENCE: 064335.0191
CURRENT APPLICATION NUMBER: US/11/072,512
CURRENT FILING DATE: 2005-03-07
PRIOR APPLICATION NUMBER: US 60/350,978
PRIOR FILING DATE: 2002-01-25
PRIOR APPLICATION NUMBER: JP 2001-379298
PRIOR FILING DATE: 2001-11-05
NUMBER OF SEQ ID NOS: 4096
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2976
LENGTH: 331
TYPE: PRT
ORGANISM: Homo sapiens
US-11-072-512-2976

```

```

QY      1 CLXXXTREIXIVE 12
|:|:|:|:|:|:|:|:|:|:|
Db      32 CMEDIDFIRIVE 43

Query Match Similarity 77.8%; Score 35; DB 11; Length 331;
Best Local Similarity 50.0%; Pred. No. 62;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

```

RESULT 4
US-11-072-512-3390
Sequence 3390, Application US/11072512
Publication No. US20060029945A1
GENERAL INFORMATION:
APPLICANT: ISOGAI, TAKAO
APPLICANT: SUGIYAMA, TOMOYASU
APPLICANT: OTSUKI, TETSUJI
APPLICANT: WAKAMATSU, AI
APPLICANT: SATO, HIROYUKI
APPLICANT: ISHII, SHIZUKO
APPLICANT: YAMAMOTO, JUN-ICHI
APPLICANT: ISONO, YUUKO
APPLICANT: HIO, YURI
APPLICANT: OTSUKA, KAORU
APPLICANT: NAGAI, KEIICHI
APPLICANT: IRIE, RYOTARO
APPLICANT: TAMECHIKA, ICHIRO
APPLICANT: SEKI, NAOHICO
APPLICANT: YOSHIKAWA, TSUTOMU
APPLICANT: OTSUKA, MOTOYUKI
APPLICANT: NAGAHARI, KENJI

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: May 4, 2006, 03:32:22 ; Search time 36.6667 Seconds
(without alignments)
33.822 Million cell updates/sec

Title: US-10-799-326-78

Perfect score: 58

Sequence: 1 DGXIRATXXHFXMT 15

Scoring table: BLOSUM62DX

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 100 summaries

Database :

Issued Patents AA: *
1: /cgn2_6/ptodata/1/1aa/5_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/6_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/H_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/PCUTS_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/RE_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	58	100.0	124 2 US-09-800-170-10	Sequence 10, Appl
2	46	79.3	610 2 US-09-800-170-22	Sequence 22, Appl
3	45	77.6	95 2 US-08-946-329A-78	Sequence 78, Appl
4	45	77.6	95 2 US-09-562-914-78	Sequence 78, Appl
5	45	77.6	135 2 US-09-800-170-54	Sequence 54, Appl
6	45	77.6	198 2 US-09-800-170-25	Sequence 25, Appl
7	45	77.6	199 2 US-09-800-170-28	Sequence 28, Appl
8	45	77.6	431 2 US-09-800-170-1	Sequence 1, Appl
9	45	77.6	434 2 US-09-800-170-48	Sequence 48, Appl
10	45	77.6	437 2 US-09-800-170-49	Sequence 49, Appl
11	45	77.6	438 2 US-09-800-170-30	Sequence 30, Appl
12	45	77.6	438 2 US-09-800-170-32	Sequence 32, Appl
13	45	77.6	438 2 US-09-800-170-34	Sequence 34, Appl
14	45	77.6	438 2 US-09-800-170-36	Sequence 36, Appl
15	45	77.6	438 2 US-09-800-170-38	Sequence 38, Appl
16	45	77.6	438 2 US-09-800-170-40	Sequence 40, Appl
17	45	77.6	438 2 US-09-800-170-42	Sequence 42, Appl
18	45	77.6	438 2 US-09-800-170-44	Sequence 44, Appl
19	45	77.6	438 2 US-08-946-329A-76	Sequence 76, Appl
20	42	72.4	106 2 US-08-946-329A-76	Sequence 76, Appl
21	42	72.4	106 2 US-09-562-914-76	Sequence 76, Appl
22	41	70.7	107 2 US-08-946-329A-85	Sequence 85, Appl
23	41	70.7	107 2 US-09-562-914-85	Sequence 85, Appl
24	40	69.0	152 2 US-09-583-110-4814	Sequence 4814, Ap
25	40	69.0	209 2 US-09-107-433-3477	Sequence 3477, Ap
26	40	69.0	811 2 US-08-983-275-2	Sequence 2, Appl
27	39	67.2	147 2 US-09-800-170-11	Sequence 11, Appl

28	39	67.2	379 2 US-09-492-709A-355	Sequence 355, App
29	39	67.2	911 2 US-09-902-540-12141	Sequence 12141, A
30	38	65.5	94 2 US-08-946-329A-88	Sequence 88, Appl
31	38	65.5	94 2 US-09-562-914-88	Sequence 88, Appl
32	38	65.5	272 2 US-09-069-023-24	Sequence 24, Appl
33	38	65.5	280 2 US-09-328-352-5799	Sequence 5799, Ap
34	38	65.5	389 2 US-09-489-039A-8540	Sequence 8540, Ap
35	38	65.5	909 2 US-09-252-991A-30503	Sequence 30503, A
36	37	63.8	117 1 US-08-901-306-5	Sequence 5, Appl
37	37	63.8	183 2 US-09-198-452A-612	Sequence 612, App
38	37	63.8	274 2 US-09-902-540-11769	Sequence 11769, A
39	37	63.8	275 2 US-09-543-661A-7448	Sequence 7448, Ap
40	37	63.8	292 2 US-09-328-352-4997	Sequence 4997, Ap
41	37	63.8	317 1 US-08-973-275-1	Sequence 1, Appl
42	37	63.8	347 2 US-09-097-889-14	Sequence 14, Appl
43	37	63.8	347 2 US-09-098-079-14	Sequence 14, Appl
44	37	63.8	358 2 US-09-328-352-7579	Sequence 7579, Ap
45	37	63.8	508 2 US-09-332-552E-4	Sequence 4, Appl
46	37	63.8	544 2 US-09-248-786A-14307	Sequence 14307, A
47	37	63.8	756 2 US-09-438-185A-574	Sequence 574, App
48	37	63.8	1479 2 US-08-840-062-2	Sequence 2, Appl
49	37	63.8	1479 2 US-08-840-062-2	Sequence 4, Appl
50	36	62.1	54 2 US-09-621-976-4577	Sequence 4577, Ap
51	36	62.1	89 2 US-09-489-039A-8970	Sequence 8970, Ap
52	36	62.1	144 2 US-10-104-047-3457	Sequence 3457, Ap
53	36	62.1	148 2 US-09-252-991A-21337	Sequence 21337, Ap
54	36	62.1	180 2 US-09-270-767-58131	Sequence 58131, A
55	36	62.1	205 2 US-09-949-016-9316	Sequence 9316, Ap
56	36	62.1	246 2 US-09-902-540-13920	Sequence 13920, A
57	36	62.1	269 1 US-07-857-224B-79	Sequence 79, Appl
58	36	62.1	304 1 US-08-701-191A-27	Sequence 27, Appl
59	36	62.1	304 2 US-09-664-526-27	Sequence 27, Appl
60	36	62.1	314 2 US-09-252-991A-20590	Sequence 20590, A
61	36	62.1	316 1 US-08-278-089A-16	Sequence 16, Appl
62	36	62.1	316 2 US-08-838-957A-15	Sequence 15, Appl
63	36	62.1	371 2 US-09-303-518D-328	Sequence 328, App
64	36	62.1	373 2 US-09-252-991A-21230	Sequence 21230, A
65	36	62.1	379 2 US-09-583-110-3811	Sequence 3811, Ap
66	36	62.1	382 2 US-09-107-433-3703	Sequence 3703, Ap
67	36	62.1	423 2 US-09-270-767-42815	Sequence 42815, A
68	36	62.1	432 2 US-09-949-016-6747	Sequence 6747, Ap
69	36	62.1	434 2 US-09-303-518D-334	Sequence 334, App
70	36	62.1	458 2 US-09-489-039A-10215	Sequence 10215, A
71	36	62.1	459 2 US-09-328-352-7963	Sequence 7963, Ap
72	36	62.1	465 2 US-09-538-092-877	Sequence 877, App
73	36	62.1	466 2 US-09-610-401-3	Sequence 3, Appl
74	36	62.1	466 2 US-09-610-401-4	Sequence 4, Appl
75	36	62.1	466 2 US-09-167-306-12	Sequence 12, Appl
76	36	62.1	466 2 US-09-949-016-6351	Sequence 6351, App
77	36	62.1	466 2 US-10-012-231A-267	Sequence 267, App
78	36	62.1	466 2 US-10-015-389A-267	Sequence 267, App
79	36	62.1	466 2 US-10-006-768A-267	Sequence 267, App
80	36	62.1	466 2 US-10-015-671A-267	Sequence 267, App
81	36	62.1	466 2 US-10-015-393A-267	Sequence 267, App
82	36	62.1	466 2 US-10-011-833A-267	Sequence 267, App
83	36	62.1	466 2 US-10-006-041A-267	Sequence 267, App
84	36	62.1	466 2 US-10-012-064A-267	Sequence 267, App
85	36	62.1	466 2 US-09-543-681A-6444	Sequence 6444, Ap
86	36	62.1	506 2 US-09-949-016-7650	Sequence 7650, Ap
87	36	62.1	534 2 US-09-198-672-2	Sequence 2, Appl
88	36	62.1	534 2 US-09-199-290-9	Sequence 9, Appl
89	36	62.1	534 2 US-09-632-392-2	Sequence 9, Appl
90	36	62.1	534 2 US-09-455-679-1	Sequence 1, Appl
91	36	62.1	534 2 US-09-351-814-2	Sequence 2, Appl
92	36	62.1	534 2 US-09-821-616-9	Sequence 9, Appl
93	36	62.1	544 2 US-09-489-039A-10622	Sequence 10622, A
94	36	62.1	616 2 US-09-236-063-1	Sequence 1, Appl
95	36	62.1	640 2 US-09-351-814-13	Sequence 13, Appl
96	36	62.1	642 2 US-09-252-991A-23588	Sequence 23588, A
97	36	62.1	671 2 US-09-303-518D-330	Sequence 330, App
98	36	62.1	671 2 US-09-303-518D-332	Sequence 332, App
99	36	62.1	671 2 US-09-303-518D-336	Sequence 336, App
100	36	62.1	971 2 US-09-902-540-12576	Sequence 12576, A

ALIGNMENTS

```
RESULT 1
US-09-800-170-10
; Sequence 10, Application US/09800170
; Patent No. 6481667
; GENERAL INFORMATION:
; APPLICANT: Kinseila, Todd
; TITLE OF INVENTION: IN VIVO PRODUCTION OF CYCLIC PEPTIDES
; FILE REFERENCE: A-68614-1/DJB/RMS/RMK
; CURRENT APPLICATION NUMBER: US/09/800,170
; CURRENT FILING DATE: 2001-08-28
; PRIOR APPLICATION NUMBER: US 60/187,130
; PRIOR FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 10
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Synechocystis PCC6803
US-09-800-170-10

Query Match      100.0%; Score 58; DB 2; Length 124;
Best Local Similarity 66.7%; Pred. No. 0.03;
Matches 10; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY      1 DGXVIRATXXHXFXT 15
Db      63 DGXVIRATSDHRLFT 77

RESULT 2
US-09-800-170-22
; Sequence 22, Application US/09800170
; Patent No. 6481667
; GENERAL INFORMATION:
; APPLICANT: Kinseila, Todd
; TITLE OF INVENTION: IN VIVO PRODUCTION OF CYCLIC PEPTIDES
; FILE REFERENCE: A-68614-1/DJB/RMS/RMK
; CURRENT APPLICATION NUMBER: US/09/800,170
; CURRENT FILING DATE: 2001-08-28
; PRIOR APPLICATION NUMBER: US 60/187,130
; PRIOR FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 22
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Pyrococcus abyssi
US-09-800-170-22

Query Match      79.3%; Score 46; DB 2; Length 610;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 7; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY      2 GXVIRATXXHXFXT 15
Db      66 GRIEATGDHMLFT 79

RESULT 3
US-08-946-329A-78
; Sequence 78, Application US/08946329A
; Patent No. 6057091
; GENERAL INFORMATION:
; APPLICANT: Beachy, Philip A.
; APPLICANT: Porter, Jeffrey A.
; TITLE OF INVENTION: NOVEL HEDGEHOG-DERIVED POLYPEPTIDES
; NUMBER OF SEQUENCES: 109
; CORRESPONDENCE ADDRESS:
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FASTSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/562,914
; FILING DATE: 02-May-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/946,329
```

```
ADDRESSER: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/946,329A
FILING DATE: 07-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/061,323
FILING DATE: 07-OCT-1996
APPLICATION NUMBER: 08/729,743
FILING DATE: 10-JUL-1996
APPLICATION NUMBER: 08/567,357
FILING DATE: 04-DEC-1995
APPLICATION NUMBER: 08/349,498
FILING DATE: 02-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Haile, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07265/140001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 78:
SEQUENCE CHARACTERISTICS:
LENGTH: 95 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-946-329A-78

Query Match      77.6%; Score 45; DB 2; Length 95;
Best Local Similarity 50.0%; Pred. No. 3.6;
Matches 7; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY      2 GXVIRATXXHXFXT 15
Db      27 GRIKATANHRLFT 40

RESULT 4
US-09-562-914-78
; Sequence 78, Application US/09562914
; Patent No. 6911528
; GENERAL INFORMATION:
; APPLICANT: Beachy, Philip A.
; APPLICANT: Porter, Jeffrey A.
; TITLE OF INVENTION: NOVEL HEDGEHOG-DERIVED POLYPEPTIDES
; NUMBER OF SEQUENCES: 109
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FASTSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/562,914
; FILING DATE: 02-May-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/946,329
```


GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: May 4, 2006, 03:33:07 ; Search time 106.667 Seconds
(without alignments)
47.006 Million cell updates/sec

Title: US-10-799-326-77
Perfect score: 45
Sequence: 1 CLXXXTREXIVE 12

Scoring table: BIOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues
Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Published Applications AA Main:
1: /cgn2_6/ptodaca/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodaca/1/pubpaa/US08_PUBCOMB.pep:*
3: /cgn2_6/ptodaca/1/pubpaa/US09_PUBCOMB.pep:*
4: /cgn2_6/ptodaca/1/pubpaa/US10A_PUBCOMB.pep:*
5: /cgn2_6/ptodaca/1/pubpaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/ptodaca/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Match	Query Length	DB ID	Description
1	45	100.0	12	4	US-10-799-326-73
2	45	100.0	12	4	US-10-799-326-77
3	45	100.0	13	4	US-10-799-326-75
4	45	100.0	78	4	US-10-799-326-71
5	45	100.0	123	4	US-10-356-088-1
6	45	100.0	123	4	US-10-356-088-23
7	45	100.0	123	4	US-10-353-454-1
8	45	100.0	123	4	US-10-353-454-23
9	45	100.0	123	4	US-10-799-326-1
10	45	100.0	123	4	US-10-799-326-23
11	45	100.0	124	3	US-09-800-770-10
12	45	100.0	124	6	US-10-422-536-10
13	45	100.0	124	6	US-11-011-880-10
14	45	100.0	326	4	US-10-356-088-38
15	45	100.0	326	4	US-10-799-326-38
16	45	100.0	416	5	US-10-833-851-25
17	45	100.0	461	5	US-10-833-851-4
18	40	88.9	817	6	US-11-097-143-7896
19	39	86.7	399	4	US-10-081-816-64
20	39	86.7	399	4	US-10-447-328-62
21	39	86.7	426	4	US-10-282-122A-78185
22	39	86.7	426	5	US-10-496-207-8
23	39	86.7	426	5	US-10-732-923-11590
24	39	86.7	427	4	US-10-282-122A-68705
25	39	86.7	914	6	US-11-097-143-31677
26	36	80.0	62	4	US-10-424-599-216407
27	36	80.0	112	5	US-10-732-923-11768

28	36	80.0	325	4	US-10-425-115-262616	Sequence 262616,
29	36	80.0	328	4	US-10-424-599-208315	Sequence 208315,
30	36	80.0	426	4	US-10-369-493-7952	Sequence 7952, Ap
31	36	80.0	427	4	US-10-369-493-394	Sequence 394, App
32	36	80.0	469	5	US-10-732-923-11420	Sequence 11420, A
33	36	80.0	479	4	US-10-369-493-19048	Sequence 19048, A
34	36	80.0	526	4	US-10-282-122A-54914	Sequence 54914, A
35	36	80.0	553	4	US-10-289-762-1100	Sequence 1100, Ap
36	36	80.0	714	4	US-10-369-493-10049	Sequence 10049, A
37	36	80.0	783	4	US-10-369-493-20178	Sequence 20178, A
38	36	80.0	832	4	US-10-437-963-141613	Sequence 141613,
39	36	80.0	1188	4	US-10-282-122A-53497	Sequence 53497, A
40	36	80.0	7746	4	US-10-156-761-7965	Sequence 7965, Ap
41	35	77.8	41	4	US-10-424-599-193757	Sequence 193757,
42	35	77.8	62	3	US-09-864-761-37476	Sequence 37476, A
43	35	77.8	62	3	US-09-864-761-48761	Sequence 48761, A
44	35	77.8	95	5	US-10-732-923-11430	Sequence 11430, A
45	35	77.8	110	4	US-10-029-386-34140	Sequence 34140, A
46	35	77.8	112	5	US-10-732-923-11498	Sequence 11498, A
47	35	77.8	112	5	US-10-732-923-11505	Sequence 11505, A
48	35	77.8	112	5	US-10-732-923-11528	Sequence 11528, A
49	35	77.8	112	5	US-10-732-923-11531	Sequence 11531, A
50	35	77.8	112	5	US-10-732-923-11830	Sequence 11830, A
51	35	77.8	188	4	US-10-424-599-167924	Sequence 167924,
52	35	77.8	188	4	US-10-425-115-244049	Sequence 244049,
53	35	77.8	206	5	US-10-732-923-11642	Sequence 11642, A
54	35	77.8	239	5	US-10-488-197-69	Sequence 69, Appl
55	35	77.8	273	4	US-10-156-761-14332	Sequence 14332, A
56	35	77.8	309	5	US-10-732-923-11687	Sequence 11687, A
57	35	77.8	321	5	US-10-732-923-11739	Sequence 11739, A
58	35	77.8	322	5	US-10-732-923-11641	Sequence 11641, A
59	35	77.8	331	4	US-10-104-047-2976	Sequence 2976, Ap
60	35	77.8	331	5	US-10-732-923-11858	Sequence 11858, A
61	35	77.8	418	4	US-10-094-749-3072	Sequence 3072, Ap
62	35	77.8	418	4	US-10-408-765A-2512	Sequence 2512, Ap
63	35	77.8	427	4	US-10-369-493-21187	Sequence 21187, A
64	35	77.8	431	5	US-10-732-923-11686	Sequence 11686, A
65	35	77.8	455	4	US-10-425-115-288959	Sequence 288959,
66	35	77.8	558	5	US-10-450-763-48215	Sequence 48215, A
67	35	77.8	591	5	US-10-488-197-71	Sequence 71, Appl
68	35	77.8	748	3	US-09-815-242-130278	Sequence 130278, A
69	35	77.8	748	3	US-09-815-242-138974	Sequence 138974, A
70	35	77.8	748	4	US-10-369-493-23534	Sequence 23534, A
71	35	77.8	748	4	US-10-369-493-23534	Sequence 23534, A
72	35	77.8	815	4	US-10-104-047-33390	Sequence 76203, A
73	34	75.6	24	4	US-10-283-940-38	Sequence 38, Appl
74	34	75.6	24	6	US-11-022-454-38	Sequence 38, Appl
75	34	75.6	36	4	US-10-424-599-241693	Sequence 241693,
76	34	75.6	60	4	US-10-450-763-30475	Sequence 30475, A
77	34	75.6	60	4	US-10-437-963-133524	Sequence 133524,
78	34	75.6	61	3	US-09-764-891-3078	Sequence 3078, Ap
79	34	75.6	78	4	US-10-425-115-186663	Sequence 186663,
80	34	75.6	102	3	US-09-746-783-44	Sequence 44, Appl
81	34	75.6	107	5	US-10-732-923-11471	Sequence 11471, A
82	34	75.6	112	5	US-10-732-923-11406	Sequence 11406, A
83	34	75.6	112	5	US-10-732-923-11530	Sequence 11530, A
84	34	75.6	121	4	US-10-424-599-231335	Sequence 231335,
85	34	75.6	127	4	US-10-425-114-72380	Sequence 72380, A
86	34	75.6	131	4	US-10-425-115-337830	Sequence 337830,
87	34	75.6	150	3	US-09-781-867-4	Sequence 4, Appl
88	34	75.6	226	3	US-09-810-151-6	Sequence 6, Appl
89	34	75.6	226	3	US-09-961-721-6	Sequence 6, Appl
90	34	75.6	226	4	US-10-045-367A-5	Sequence 5, Appl
91	34	75.6	226	4	US-10-170-789-47	Sequence 47, Appl
92	34	75.6	226	4	US-10-170-789-55	Sequence 55, Appl
93	34	75.6	237	4	US-10-203-064A-8	Sequence 8, Appl
94	34	75.6	248	4	US-10-282-122A-63043	Sequence 63043, A
95	34	75.6	260	4	US-10-369-493-11122	Sequence 1122, Ap
96	34	75.6	271	4	US-10-266-035-4	Sequence 4, Appl
97	34	75.6	271	4	US-10-391-364-23	Sequence 23, Appl
98	34	75.6	297	4	US-10-312-273-229	Sequence 29, App
99	34	75.6	319	5	US-10-732-923-11640	Sequence 11640, A
100	34	75.6	324	4	US-10-369-493-937	Sequence 937, App

ALIGNMENTS

```
RESULT 1
US-10-799-326-73
; Sequence 73, Application US/10799326
; Publication No. US20040172688A1
; GENERAL INFORMATION:
; APPLICANT: E.I. DuPont de Nemours, & Company
; APPLICANT: Yadav, Narendra S.
; APPLICANT: Yang, Jianjun Gene
; TITLE OF INVENTION: Intein-Mediated Protein Splicing
; FILE REFERENCE: CL1806 US CIP
; CURRENT FILING DATE: 2004-03-12
; PRIOR APPLICATION NUMBER: US 60/354395
; PRIOR FILING DATE: 2002-02-04
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 73
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Synechocystis sp. PCC6803
US-10-799-326-73

Query Match      100.0%; Score 45; DB 4; Length 12;
Best Local Similarity 66.7%; Pred. No. 0.36;
Matches 8; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY      1 CLXXXTEIXTVE 12
      ||::||::||
Db      1 CLSFGTEILTVE 12

RESULT 2
US-10-799-326-77
; Sequence 77, Application US/10799326
; Publication No. US20040172688A1
; GENERAL INFORMATION:
; APPLICANT: E.I. DuPont de Nemours, & Company
; APPLICANT: Yadav, Narendra S.
; APPLICANT: Yang, Jianjun Gene
; TITLE OF INVENTION: Intein-Mediated Protein Splicing
; FILE REFERENCE: CL1806 US CIP
; CURRENT FILING DATE: 2004-03-12
; PRIOR APPLICATION NUMBER: US 60/354395
; PRIOR FILING DATE: 2002-02-04
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 77
; LENGTH: 12
; TYPE: PRT
; ORGANISM: artificial
; FEATURE:
; OTHER INFORMATION: compiled sequence
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (3)..(5)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
; NAME/KEY: misc.feature
; LOCATION: (9)..(9)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-10-799-326-77

Query Match      100.0%; Score 45; DB 4; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.36;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CLXXXTEIXTVE 12
```

```
Db      1 |||||
      1 CLXXXTEIXTVE 12

RESULT 3
US-10-799-326-75
; Sequence 75, Application US/10799326
; Publication No. US20040172688A1
; GENERAL INFORMATION:
; APPLICANT: E.I. DuPont de Nemours, & Company
; APPLICANT: Yadav, Narendra S.
; APPLICANT: Yang, Jianjun Gene
; TITLE OF INVENTION: Intein-Mediated Protein Splicing
; FILE REFERENCE: CL1806 US CIP
; CURRENT FILING DATE: 2004-03-12
; PRIOR APPLICATION NUMBER: US 60/354395
; PRIOR FILING DATE: 2002-02-04
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 75
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Trichodesmium erythraeum
US-10-799-326-75

Query Match      100.0%; Score 45; DB 4; Length 13;
Best Local Similarity 66.7%; Pred. No. 0.39;
Matches 8; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY      1 CLXXXTEIXTVE 12
      ||::||::||
Db      1 CLVTEILTVE 12

RESULT 4
US-10-799-326-71
; Sequence 71, Application US/10799326
; Publication No. US20040172688A1
; GENERAL INFORMATION:
; APPLICANT: E.I. DuPont de Nemours, & Company
; APPLICANT: Yadav, Narendra S.
; APPLICANT: Yang, Jianjun Gene
; TITLE OF INVENTION: Intein-Mediated Protein Splicing
; FILE REFERENCE: CL1806 US CIP
; CURRENT FILING DATE: 2004-03-12
; PRIOR APPLICATION NUMBER: US 60/354395
; PRIOR FILING DATE: 2002-02-04
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 71
; LENGTH: 78
; TYPE: PRT
; ORGANISM: Synechocystis sp. PCC6803
US-10-799-326-71

Query Match      100.0%; Score 45; DB 4; Length 78;
Best Local Similarity 66.7%; Pred. No. 3.1;
Matches 8; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY      1 CLXXXTEIXTVE 12
      ||::||::||
Db      1 CLSFGTEILTVE 12

RESULT 5
US-10-336-088-1
; Sequence 1, Application US/10356088
; Publication No. US20030167533A1
; GENERAL INFORMATION:
; APPLICANT: Yadav, Narendra S.
; APPLICANT: Yang, Jianjun Gene
```